SANDIA REPORT

SAND2010-0388
Unlimited Release
Printed February 2010

Description of the Sandia National Laboratories Science, Technology & Engineering Metrics Process

Gretchen Jordan, Pete Oelschlaeger, Alan Burns, Randy Watkins, and Tim Trucano

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy under Contract DE-AC04-94AL8500.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2010-0388
Unlimited Release
Printed February 2010

Description of the Sandia National Laboratories Science, Technology & Engineering Metrics Process

Authors:

Gretchen Jordan, Pete Oelschlaeger, Alan Burns, Randy Watkins, ST&E Strategy and University Research Development Department, and Tim Trucano, Optimization and Uncertainty Estimation Department

P.O. Box 5800,
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

There has been a concerted effort since 2007 to establish a dashboard of metrics for the Science, Technology, and Engineering (ST&E) work at Sandia National Laboratories. These metrics are to provide a self assessment mechanism for the ST&E Strategic Management Unit (SMU) to complement external expert review and advice and various internal self assessment processes. The data and analysis will help ST&E Managers plan, implement, and track strategies and work in order to support the critical success factors of nurturing core science and enabling laboratory missions. The purpose of this SAND report is to provide a guide for those who want to understand the ST&E SMU metrics process. This report provides an overview of why the ST&E SMU wants a dashboard of metrics, some background on metrics for ST&E programs from existing literature and past Sandia metrics efforts, a summary of work completed to date, specifics on the portfolio of metrics that have been chosen and the implementation process that has been followed, and plans for the coming year to improve the ST&E SMU metrics process.

Acknowledgments

The authors want to acknowledge the leadership of Richard Stulen, then Chief Technology Officer, in initiating and nurturing this effort. Also, the members of the Sandia Science Advisory Board (SSAB) were instrumental in encouraging and reviewing the development of these metrics for Science, Technology and Engineering work at Sandia National Laboratories. Initial planning was enriched by the larger metrics advisory group which included Wendy Cieslak, Ron Hartwig, Laura McNamara, Hank Westrich, Jeff Brinker, and Carl Peterson, in addition to the authors of this report.

This page intentionally left blank

Executive Summary

The pressure to have more quantitative performance measures for Science, Technology, and Engineering (ST&E) had been growing and in late Fall 2006 then Chief Technology Officer Rick Stulen established an advisory group of Sandia senior scientists and social scientists with expertise in ST&E metrics and organizational change and charged the group with defining metrics and an implementation plan and "to get this right." Specifically, the ST&E metrics group was to (1) define a portfolio of metrics that help Sandia ST&E meet its strategic goals and (2) increase Sandia comprehension and appreciation of the creation and value of ST&E at the laboratories. The group was also to understand the importance of changing behavior, understand the data issues (metrics reflect values, what/who/how to measure, application of the data), connect to Laboratory Directed R&D (LDRD) measurement for input, and practice due diligence with respect to other related initiatives.

The scope of this effort was limited to the major corporate goal ST&E supports, including LDRD, which is "Create breakthrough results through science and engineering." The plan was to examine best practices, speak with focus groups, and have a metrics plan to the Sandia Science Advisory Board (SSAB) in 2007. A subset of the advisory group became the metrics working group after plans were set.

The metrics working group reviewed existing literature and their Sandia experiences, and drafted a proposal for a portfolio of metrics using a balanced scorecard, a comprehensive approach to choosing and displaying metrics that is familiar to Sandians. The group drafted guiding principles and a preliminary phased implementation plan. This was presented to focus groups in Albuquerque and California between February and June 2007. That input slightly modified the proposed small set of metrics, and the resulting set is shown in Figure ES 1.

The left hand side of the dashboard of metrics measures "What" ST&E does, from capabilities to outputs such as publications and technical advances, and the value these outputs provide to customers who include internal Sandia mission areas, DOE and other federal agencies, and the ST&E community. The right hand side measures "How" that ST&E gets done. Just as a hospital must simultaneously give its patients excellent care and control costs, ST&E efforts produce quality, relevant breakthrough results (the "What") through strategic planning and investments, excellent research management that drives innovation, and through collaborative work within ST&E and with partners and customers. The resulting six areas in the dashboard form an integrated portfolio of metrics. Achieving one metric is dependent on achieving several other metrics.

STE SMU and (selected) LDRD Metrics	
ST&E Value and Impact	**Strategic Collaborations**
ST&E Leadership and Stewardship World citation factor Professional society positions, awards Invited talks Visitors to SNL (domestic and foreign) LDRD Citations **Mission impact** Mission related milestones met Accomplishments	**Technical** Co-authored papers Presentations and Conference attendance University collaborations/contracts **Industry** CRADAs: number and funding
Technical Excellence	**Research Management Excellence**
Science Advances Journal publications SAND reports DOE annual ST&E rating LDRD Publications **Technical Advances** Intellectual Property R&D 100 Awards LDRD Intellectual Property LDRD R&D 100 Awards	**Work Environment** Health of the research environment Benefits of RF External Review **External Reviews** Quality of external panels and reviewers Recommendations acted upon External customer satisfaction
ST&E Capabilities	**ST&E Strategy**
Staff New hires Staff Retention **Technology Infrastructure & Facilities** Capital Equipment Investment Utilization by SMG in Computing, MESA, Environmental Test	**Portfolio** ST&E funds level Mix of Basic-Applied-Development LDRD mix of Discover-Create-Prove **Planning** ST&E and RF strategic plans completed ST&E O/G/M Milestones met

Inputs → Outputs → Value to Customers

WHAT we do ← HOW we do it

Early in the implementation phase it was decided that summary data was needed as soon as possible, so these metrics have been collected for the ST&E SMU organization as a whole, with differences shown for Research Foundations where feasible and useful.

The Sandia Science Advisory Board (SSAB) has been instrumental in encouraging and reviewing the development of these metrics for the ST&E SMU. A two slide summary of the resulting small set of metrics was presented to the SSAB in September 2007. Work then began in earnest to populate the chosen set of metrics and show this set and the implementation plan to the SSAB in March 2008. This comprehensive portfolio of metrics information was also introduced to the ST&E Council and to the Laboratory Leadership Team. Response was very positive. Additional metrics data were presented to the SSAB in September 2008 and March 2009. Portions of the ST&E SMU metrics, those by Research Foundation, have also been used during RF External Review Panel meetings since 2008.

Trend data is being collected where this is possible. Few benchmarks are available, although there is hope that agreements could be made with sister laboratories to develop these. Currently, targets are either "maintain steady state" or "increase from previous years." Status categories of "Meets Expectations," "Concern-Watch," "Requires Management Attention," are derived based on the current year as a percentage of previous years, adjusting that percentage by the amount of variation in the historical data. "Quad" charts have been developed for some of the metrics to summarize why management sees the metrics trend to be an issue, and to describe decisions made on action to be taken moving forward.

The first Annual Metrics Report was completed in March 2009, in the form of a simple dashboard where it is possible to drill down to the metrics that make up those larger categories, such as "Value and Impact" and "Technical Excellence." The dashboard and metrics are posted on the ST&E SMU web site and updated quarterly with new information. An operating plan has been written to complete the data collection on an annual cycle, standardize data collection and storage, and organize presentation and dissemination of the metrics data. There will be a plan for quality assurance of the metrics data and analysis, and for reviewing the efficacy of the metrics effort in order to continuously improve it.

The metrics have been incorporated into the Laboratories' reporting system, Integrated Laboratory Management System (ILMS), and the yet-to-be-completed PerformanceSoft corporate system. The chosen metrics have been incorporated into the evidence requested in the annual DOE Laboratory Appraisal system, the Performance Evaluation Plan (PEP) and Self Assessment (PEAR).

There is a vision of the journey to an optimized and sustainable metrics system for the ST&E SMU. Maturation of performance metrics has already proceeded from ad hoc collection for external requests to planned annual data collection and reporting. Optimally, this will move to the point where performance metrics are systematic and applied for improvement. Current issues identified for improvement are:

- Utilization of ST&E metrics (setting benchmarks and targets, management use and role in driving behaviors);
- Assuring data quality;
- Adding additional metrics such as Readiness Levels of technologies, using data mining to display relevance, and better value and impact measures; and
- Improving linkages among metrics so that specific investments in capabilities can be linked to products and to value to customers.

Description of the Sandia National Laboratories ST&E Metrics Process

Table of Contents

List of Figures

List of Tables

This page intentionally left blank

Description of the Sandia National Laboratories ST&E Metrics Process

1. INTRODUCTION AND OVERVIEW

1.1 Introduction

Sandia National Laboratories (Sandia) is a multi-program laboratory operated for the U.S. Department of Energy by Sandia Corporation, a subsidiary of Lockheed Martin Corporation. Sandia's Vision is to become the laboratory that the U.S. turns to first for innovative, science-based, systems-engineering solutions to the most challenging problems that threaten peace and freedom for our nation and the globe.

The Science, Technology & Engineering Strategic Management Unit (ST&E SMU) is a cross-cutting foundational and transformational SMU. ST&E plays an active role in all three of the Laboratories' Strategic Management Groups: Nuclear Weapons, National Security Technologies and Systems, and Laboratory Transformation, as the provider of the science and technology foundations that underpin Sandia. The ST&E SMU creates, integrates and applies capabilities needed to address national security challenges through strategic investments in six research foundations (RFs): Bioscience, Computer and Information Sciences, Engineering Sciences, Materials Science & Technology, Microelectronics & Microsystems, and Pulsed Power Science.

The Vision for the ST&E SMU, as expressed in the FY2007 SMU Strategic Plan is "Securing America's future through discovery and innovation at the interface of science and engineering." The ST&E SMU strategy is guided by two principles:

1. Nurture the Core
To ensure that the fundamental science and engineering core is vibrant and pushing the forefront of knowledge.

2. Enable the Missions
To apply that science base to enable and deliver effective and innovative solutions to current needs and to anticipate future needs.

There has been a concerted effort since 2007 to establish a portfolio -- referred to as a dashboard -- of metrics, applying beyond Laboratory Directed Research and Development (LDRD) projects, to provide a self assessment mechanism for the ST&E SMU in addition to external expert review and advice and various internal self assessment processes. The primary objective of this project has been to design a core portfolio of metrics and a metrics process to collect, analyze and report on this portfolio of metrics data to help ST&E management plan, implement, and track strategies and work in order to support the critical success factors of nurturing core science and enabling laboratory missions.

The purpose of this SAND report is to provide a guide for Sandians who want to understand the Science, Technology, and Engineering Strategic Management Unit metrics process, or want to use the lessons learned by this effort to establish their own system of metrics.

This report first provides an overview of why the ST&E SMU would want to measure and how strategic measurement can align the vision and results of multiple levels of the organization. Section 2 provides some background on metrics for ST&E programs from existing literature and past Sandia efforts related to metrics, looking for best practices such as balanced scorecards (Kaplan and Norton 1996). Sections 3, 4 and 5 provide a summary of work completed to date, specifics on the portfolio of metrics that have been chosen and the implementation process that has been followed. Section 6 looks forward to plans for the coming year to improve the ST&E SMU metrics process.

1.2 Why Measure?

Metrics are measures of aspects of resources or performance, with a unit of measurement included. If something cannot be measured directly, a proxy or indicator may be chosen as the metric to represent what can't be measured. A "leading" indicator tends to change before that for which it is a proxy, such as a sore throat coming before a cold, while lagging indicators occur after the performance being measured, such as publications coming many months after a discovery or technical progress. A metric typically has a target for where you want to be, a base level of where you are now, and a time frame for achieving the target. Metrics give us information – the "what" – but we need analysis of the data to understand the "why" in order to determine an appropriate response. Understanding "why" is how we learn from the data and improve measured performance. Reliable leading indicators, if they can be defined, are particularly useful because the information provides a basis for timely action. Metrics can also help us with the "so what" questions our external stakeholders frequently ask about ST&E.

There are two major audiences for the ST&E SMU metrics, internal and external. In addition to the primary audience, internal ST&E SMU management -- the Chief Technology Officer (CTO) and the ST&E Council -- there is the audience of management and oversight external to the ST&E SMU, including Sandia and DOE, as well as the broader ST&E and customer community.

The major reasons for establishing a dashboard of metrics for the CTO and the ST&E SMU Council (which includes Center and RF Directors) are the following:

1) Improve strategic management of ST&E including tailoring of ST&E for mission needs, and allocation of resources;

2) Better understanding of discovery and innovation at Sandia and of the ST&E role and influence in the broader Sandia national security mission;

3) Using that understanding to help improve the Sandia ST&E environment; and

4) Measuring and driving ST&E achievement (e.g. toward discovery, innovation). This does not, however, include using these metrics for individual performance appraisal.

The purpose of metrics for the external audiences (Sandia DOE and Work for Others Programs, the DOE Sandia Site Office, Congress, the media) is to:

1) Document and communicate breakthrough results and other value to satisfy oversight and requirements, such as the DOE Annual Performance Report; and

2) Demonstrate value for expenditures to attract resources, talent, and establish primacy.

1.3 Measures Follow Strategy

Ideally measures follow strategy. Figure 1-1 illustrates interconnections between strategy, activities and metrics. Metrics can be important cogs to achieve our goals for ST&E. We want a portfolio of metrics that drive organizational and behavioral change, that is, that support the creation and implementation of the organizational strategy which may also be called "Critical Success Factors." Metrics including assessments of these Critical Success Factors and levels of strategy can be displayed together in a dashboard such as the popular "Balanced Scorecard" described in Section 2.3.

Upper level strategy drives lower level organizational and individual activities. Individual activities and achieving lower level organizational strategies help achieve SMU-level organizational strategies. Another way of saying this is that there needs to be clarity (line of sight) between all organizational levels for the metrics dashboard to be one that drives the desired behaviors. The ST&E SMU strategic plan influences the strategic plans of the Research Foundations, and vice versa. And measurement through metrics at multiple levels of activities, outputs, and outcomes then provides information to modify the goals and strategy as needed.

The ST&E metrics around strategy are to drive performance in desired directions. However, input from focus groups and common sense given the untested nature and incompleteness of the initial set of ST&E metrics, have led to the decision NOT to use these ST&E metrics in individual staff performance appraisal.

Two famous quotations have guided the development of the ST&E SMU metrics:

> *"What gets measured, gets done" (Deming)*
>
> *"Not everything that can be counted counts, and not everything that counts can be counted" (Einstein)*

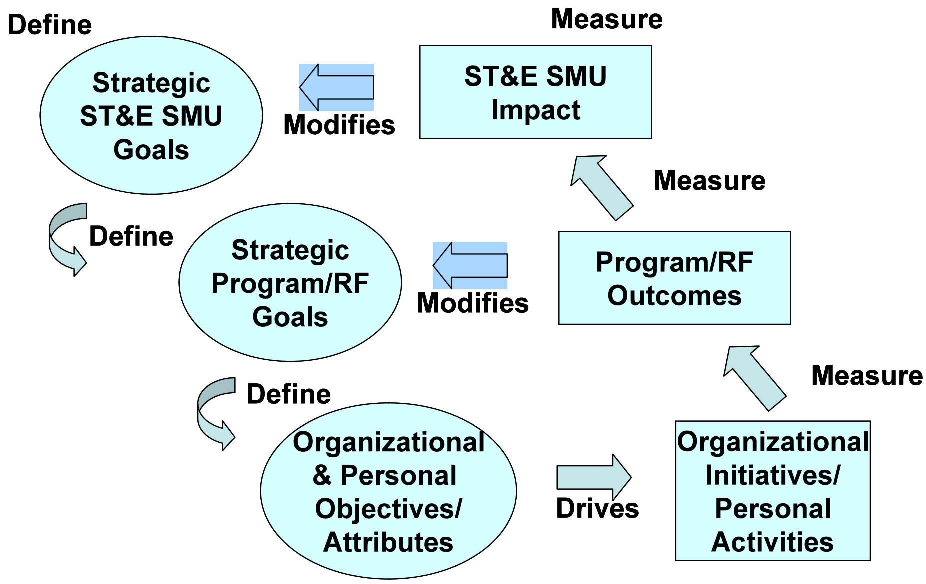


Figure 1-1. Ideally, Measures Follow Multiple Levels of Strategy

2. BACKGROUND

This section provides the background for how the ST&E SMU metrics working group arrived at the portfolio of metrics in the current dashboard of metrics, and at the design of the process for collecting, analyzing, utilizing, and improving this portfolio of metrics. The extensive literature and what Sandia and others have done on metrics are reviewed for science and technology (S&T), science and engineering, and research and development (R&D). Particular attention is paid to the balanced scorecard approach, since that is the scheme chosen for defining the ST&E SMU portfolio of metrics.

2.1 Review of the Literature

The increased attention to performance management in public R&D organizations is part of a broad trend toward increased attention to performance management in public programs generally. The Government Performance and Results Act (GPRA) passed by the U.S. Congress in 1993[1] requires strategic plans and annual performance plans and performance reports for all federal agencies. An Organization for Economic Cooperation and Development (OECD) 1997 study concluded that three public concerns were noticeable in all OECD performance management frameworks, although to different degrees: concern that governments (1) improve performance, (2) clarify responsibilities and control, and (3) realize cost savings. In 2001, the U.S. President's Management Agenda called for budget and performance integration and a new assessment process, the Program Assessment Rating Tool (PART).[2]

Managers and evaluators have found that assessing R&D performance is especially difficult because of specific qualities that are inherent to research and the scientific process. First is the open-ended nature of R&D (Rip 2003). No one can predict what discoveries will be made as a result of R&D activities because they are neither routine nor do they have specific outputs. Any goal that was predicted would most likely shift before the time to evaluate the goal had arrived. Cozzens (1999) outlines four factors that pose particular challenges for assessing the performance of R&D and meeting the requirements of GPRA:

- The attributes of research that can be tracked and measured are not always important.
- Significant research events occur unpredictably and cannot be subject to schedules.
- Many sources of funding and contribution are often integrated in a single research program.
- There is no easy, accurate method to objectively evaluate research quality or result.

Another reason that assessment of R&D is difficult is that innovative contributions to social and economic well being are typically the result of the national R&D system rather than any one component. Rip (2003) describes this as the institutional landscape of

[1] See www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html
[2] See www.whitehouse.gov/omb/part/

national R&D systems, and includes in it contributions from academic institutions, large and small public laboratories, R&D stimulation programs, and research centers. Feller (2003) also emphasizes the need for R&D performance assessment to consider collaborative research and to pay attention to the different goals and strategies of private versus public institutions.

There are several significant studies that discuss what science and technology (S&T) organizations, both government and industry, should measure and what methods for measurement are most appropriate (Brown/ARL, COSEPUP, Ellis, Geisler, NSTC). Each study has its strengths and limitations from the standpoint of the Sandia metrics benchmarking study. None of the studies provide a description of S&T metrics that is concise, current, and comprehensive (covers inputs, process, and results). Nor do any of the studies compare metrics used by government, academic, and industrial research organizations.

- The Army Research Laboratory (ARL, Brown 1997) discusses public S&T. While comprehensive with both quantitative and qualitative metrics, the ARL set does not address innovation.
- The National Research Council's Committee on Science, Engineering, and Public Policy (COSEPUP 1999) study Evaluating Federal Research Programs, contains very little detail on the metrics and methods that it recommends and its emphasis is on measuring outcomes to meet GPRA requirements.
- The Ellis book (2000) is detailed, but limited to R&D process measures used by industry.
- The Geisler book (2000) is a comprehensive review of metrics used by industry, government, and academia, although it does not compare these.
- The National Science and Technology study (1996), Assessing Fundamental Science, is very detailed on measuring outcomes of fundamental research.

International attention has been focused on assessment of publicly funded research programs. In 2005 the "World Research Evaluation Network" (WREN) was announced at an International meeting and a charter and work plan drafted. The work plan was to tackle the problems of evaluating at a systems level and building consistent quality data, including data on prospective and retrospective impacts of research. This group lost its champions and is now only informally active at the annual conference of the American Evaluation Association. [3] The considerable resource materials developed by WREN will be available on the White House Office of Science and Technology Policy Science of Science Policy website in 2010.

Considerable, mostly uncoordinated, work continues. The National Science Foundation (NSF) annually publishes science indicators, as does the OECD, but these are primarily inputs such as the amount of funds spent on R&D and the number of students graduating by discipline. The OECD developed the Frascati Manual in 1963 and supplemented this with the Oslo Manual in 2005. The OECD Science, Technology and Industry Scoreboard includes data on collaboration with public research organizations by innovating firms. The European Commission does five year assessments on each Framework Program, and

[3] See www.eval.org, the Research, Technology and Development Evaluation Topical Interest Group.

one Directorate has funded research to develop a system of metrics to supplement peer review and program impact studies. Some of Japan's research programs have developed a system of assessing programs, relying mainly on peer review and follow up studies, where performance and impact is fed back into decision making.

2.2 What Others are Doing

To the extent possible, publicly funded research has learned lessons from industry. This has been limited, however, because public research has no single bottom line like profit, but rather has very diverse missions and stakeholders. The Industrial Research Institute (IRI), a non-profit organization of over 280 leading companies that carry out over 80 percent of the industrial research effort in the United States' manufacturing sector, in the 1990s cooperatively developed the "Technology Value Pyramid" (TVP). See Figure 2-1. This is a framework for measuring the performance of industrial R&D, and for assessing the relationship between R&D programs and long-term organizational goals (Tipping et al. 1995 and Brown and Swenson 1998). The pyramid's three levels -- Foundations, Strategy, Outcomes - - are the basis for linking the key technology development and innovative processes that need to be intelligently managed and evaluated to achieve effective R&D. The key processes are represented by five managerial factors that are distributed among the three levels according to their organizational role.

A Sandia LDRD Study in 2008 looked at metrics used by several industrial firms. This study concluded that there is no single, comprehensive set of metrics for science and technology efforts used by firms. That said, "The primary metrics used at the companies studied here are the standard counting metrics of papers published, patent applications, and product introductions, as well as some measure of the overall business impact of a project on the company."

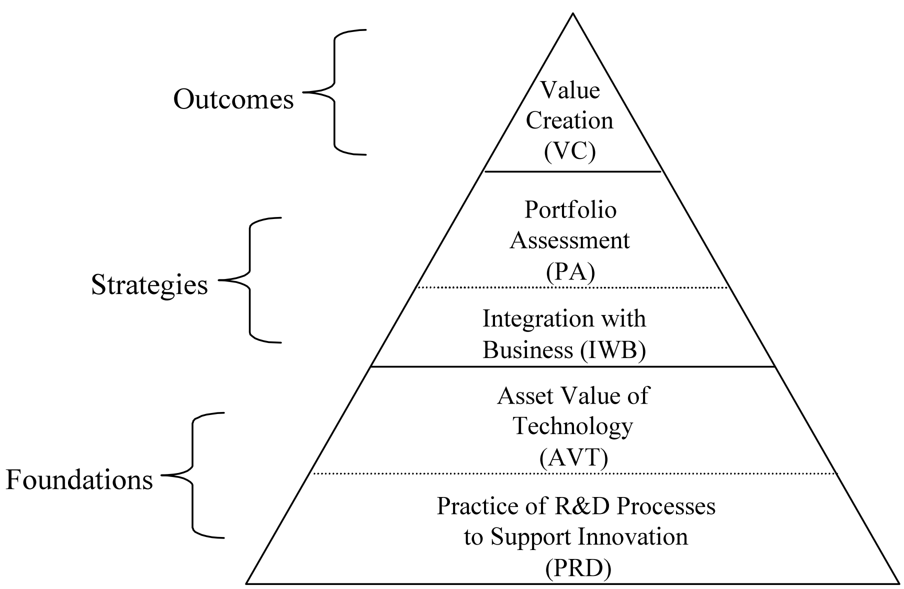


Figure 2-1. The Industrial Research Institute's Technology Value Pyramid

Our sister laboratories collect metrics common to all DOE LDRD programs. Some also have broader metrics systems. For example, the LANL Contractor Assurance Office designed, launched, and integrated a single Contractor Assurance System for all work, not just LDRD. This system includes processes and tools for managing issues, assessing and measuring key performance and Lab wide commitments. A LANL Dashboard is being used at the Director's Portfolio Review to identify emerging issues and take action to improve Laboratory performance and meet customer commitments. Summarized dashboard metrics are accessible in a LANL Briefing Book that is accessible to LANL staff. There are six focus areas measured though the dashboard: mission; science, technology and engineering; operations; business; environmental; and institutional support. Each focus area has drill-down metrics that are to be updated monthly. [4]

The Sandia ST&E SMU had developed a set of metrics in the late 1990s, but the implementation process was not institutionalized sufficiently to withstand management and staff turnover. We can learn, however, from a Metrics Benchmarking Study the ST&E SMU completed in 2001. The study was designed to address the performance measures and scorecards currently used by leading research organizations in managing science and technology efforts. The study included 24 highly regarded research organizations chosen to represent a range of federally funded laboratories and private industry research centers. A written questionnaire, telephone interviews and document exchange preceded a workshop.

Together, the survey and interview and workshop data suggested strong, common concern about better measuring and demonstrating the relevance and value of S&T, including better measures of return on investment. Other interests and concerns included metrics to assist with portfolio decisions, portfolio management, balancing short and long-term research, risk management, forward-looking portfolio valuations, and assessments for process and program improvement along with many other topics. Several individuals noted a concern that metrics, having such powerful influences, must be developed and used with caution and prudence to avert unintended consequences.

2.3 Measurement Tools – the Balanced Scorecard

The current metrics project reviewed measurement models, as well as indicators and uses discussed above. The 2001 benchmarking study found that the models most frequently used by federally funded laboratories were external peer reviews, program milestones, and balanced scorecard. For private industry most frequent were program milestones, qualitative surveys (customer/employee, etc.), six sigma, and activity based costing. Our review showed that the balanced scorecard (BSC) method is the only one of these that is a comprehensive approach to measuring performance and driving organizational strategy. It can encompass these other methods and models. Developed by Kaplan and Norton, a BSC attempts to overcome a narrow focus on financial performance (or socio-economic outcomes for public entities) by incorporating other perspectives on the performance of

[4] See http://www.lanl.gov/news/index.php/fuseaction/nb.story/story_id/%2010197.

the organization. A generic BSC showing the strategies being measured is shown in Figure 2.2.

The BSC balances out four important perspectives:

- financial performance (revenue/profit or mission)
- the customer perspective
- the business process perspective, and
- the learning and growth perspective.

Kaplan and Norton (1996, 2000) argue that the BSC is not simply a performance measurement tool, but a strategic management system. In developing metrics for all four areas, the BSC forces managers to focus on those aspects of the organization that are most important to future success. In this manner, the BSC helps managers translate strategy into operational goals, plan accordingly, measure performance in all areas and adjust strategy accordingly (Kaplan and Norton, 1996). It is a management tool that can help align smaller scale organizational objectives, with larger scale objectives in the vision and strategy, as discussed in Section 1.3 .

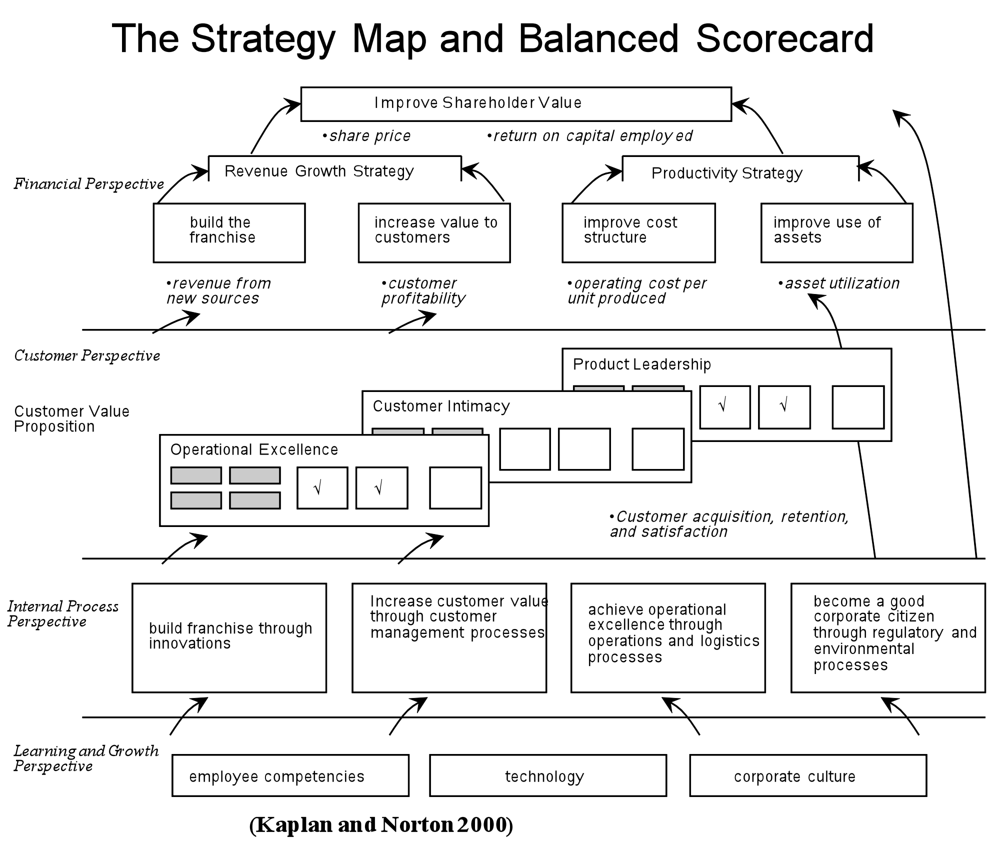


Figure 2-2. A Generic Balanced Scorecard and Strategy Map

Since its introduction, the BSC has been adopted widely across organizations and industries. While originally introduced as a tool intended for commercial organizations, the BSC has also generated interest in the public sector (Modell, 2004). With regard to R&D, including publicly-funded R&D, the BSC has found limited application, at least in the literature. Bremser and Barsky (2004) argue for a BSC approach to measuring performance in R&D that also integrates the use of the Stage Gate method.[5] However, their particular formulation of the BSC focuses on the later stages of R&D and attempts to better link R&D to avenues toward commercialization, such as the design, production and marketing units of a firm. Similarly, Eliat, Golany and Shtub (2005) develop a highly quantitative formulation of the BSC for evaluating R&D projects and portfolio selection. Interestingly, they attempt to account for interactions between R&D projects, however, the metrics utilized in the BSC focus primarily on financial aspects. Jordan et al (2006) suggest different interpretations of the BSC for science and technology programs, for example, that innovativeness, speed of innovation, and cost savings from innovation are more appropriate than research cost savings and research productivity.

In summary, the ST&E SMU metrics working group has chosen the balanced scorecard as the means to design and display a portfolio of metrics. The balanced scorecard is a way of communicating the ST&E strategy to ST&E programs and staff, and collecting metrics data to assess the implementation of that strategy. A BSC builds upon what Kaplan and Norton call a "Strategy Map." There is a quote from Alice in Wonderland, "If you don't know where you are going, it doesn't matter which road you take." Similarly, selecting a set of metrics without a map, that is, without an integrated framework for considering those measures, is not a good idea. Also note that we measure implementation of the strategy and plans. Planning can be iterative but is basically about extending the strategic mission down to programs and staff. Implementation reverses the direction, passing from staff up through programs to customers and mission. So the four perspectives of the BSC provide required line of sight as shown in Figure 1-1.

[5] For a recent article by its inventor, see Robert G. Cooper (2008).

3. SUMMARY OF WORK COMPLETED TO DATE

3.1 Overview of Work Completed

Figure 3-1 shows a high-level timeline of work completed to date, and this section provides a summary of that work: 1) the initial Chief Technology Officer (CTO) tasking, 2) initial design, and 3) implementation. More detail is provided in the sections that follow.

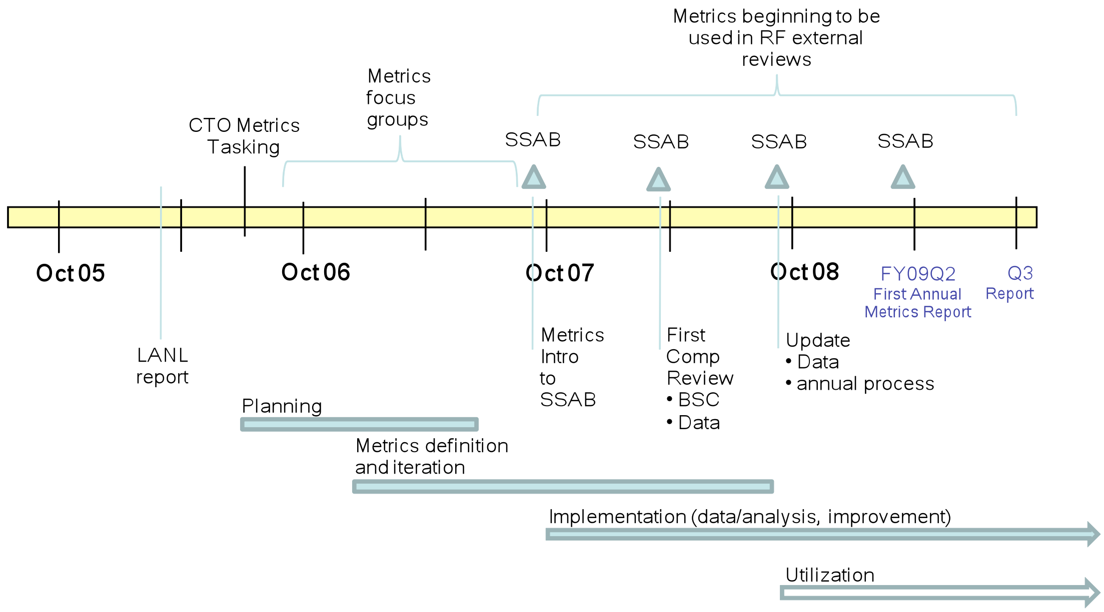


Figure 3-1. Timeline of Work Completed on ST&E SMU Metrics

3.2 The Metrics Working Group and Its Charge

The pressure to have more quantitative measures for ST&E, that is, hard credible data for evidence-based decisions, had been growing and in late Fall 2006 then-CTO Rick Stulen established an advisory group of Sandia senior scientists and social scientists with expertise in ST&E metrics and organizational change and charged the group with defining metrics and an implementation plan and "to get this right." A new internal white paper on metrics for ST&E from Los Alamos National Laboratory was part of the learning process. Specifically, the ST&E Metrics Group was to (1) define a portfolio of metrics that help Sandia ST&E meet its strategic goals and (2) increase Sandia comprehension and appreciation of the creation and value of ST&E at the laboratories. The group was also to understand the importance of changing behavior (in positive directions, recognizing other incentives in the system), understand the data issues (metrics reflect values, what/who/how to measure, application of the data), connect to

Laboratory Directed R&D (LDRD) measurement for input, and practice due diligence with respect to other initiatives.

The scope of this effort was limited to the major corporate goal ST&E supports, including LDRD, which is "Create breakthrough results through science and engineering." The plan was to examine best practices, speak with focus groups, and have a metrics plan to Sandia Science Advisory Board (SSAB) in 2007. A subset of the advisory group became the metrics working group after plans were set.

3.3 Initial Design

The metrics working group reviewed existing literature and their Sandia experiences and drafted a proposal for a portfolio of metrics using a balanced scorecard, a comprehensive approach to choosing and displaying metrics that is familiar to Sandians and described in Section 2. The group drafted guiding principles and a preliminary phased implementation plan. This was presented to focus groups in Albuquerque and California between February and June 2007. Focus groups brainstormed concerns people had about metrics, got comments on the preliminary integrative framework, and talked more specifically about "metrics," that is, data that could be collected that would indicate an objective had been achieved. Lastly (and briefly) the groups discussed the implementation of these metrics, because many concerns were related to that, not just to what would be measured.

In Appendix B the feedback from the focus group sessions is loosely summarized in three general topical areas: (1) general comments and concerns about ST&E metrics and our process; (2) specific metrics concerns; (3) alternative or modified metrics of interest. People were anxious that the metrics process be a good one. The metrics working group had already taken that into consideration, and that concern should remain a focus as the metrics process is implemented and refined. The focus group input slightly modified the proposed small set of metrics, which is shown in Appendix B. The 2009 balanced portfolio of metrics shown in a dashboard view in Figure 3.2.

3.4 Implementation

Early in the implementation phase it was decided that summary data was needed as soon as possible, so the metrics have been collected for the ST&E SMU organization as a whole, with differences shown for Research Foundations where feasible and useful.

The Sandia Science Advisory Board (SSAB) has been instrumental in encouraging and reviewing the development of these metrics for the ST&E SMU. A two slide summary of the resulting small set of metrics was presented to the SSAB in September 2007. Work then began in earnest to populate the chosen set of metrics and show this set and the implementation plan to the SSAB in March 2008. This comprehensive portfolio of metrics information was also introduced to the ST&E Council and to the Laboratory Leadership Team. Response was very positive. Additional metrics data were presented to the SSAB in September 2008 and March 2009. Portions of the ST&E SMU metrics, those by Research Foundation, were also used during RF External Review Panel meetings in 2008 and 2009.

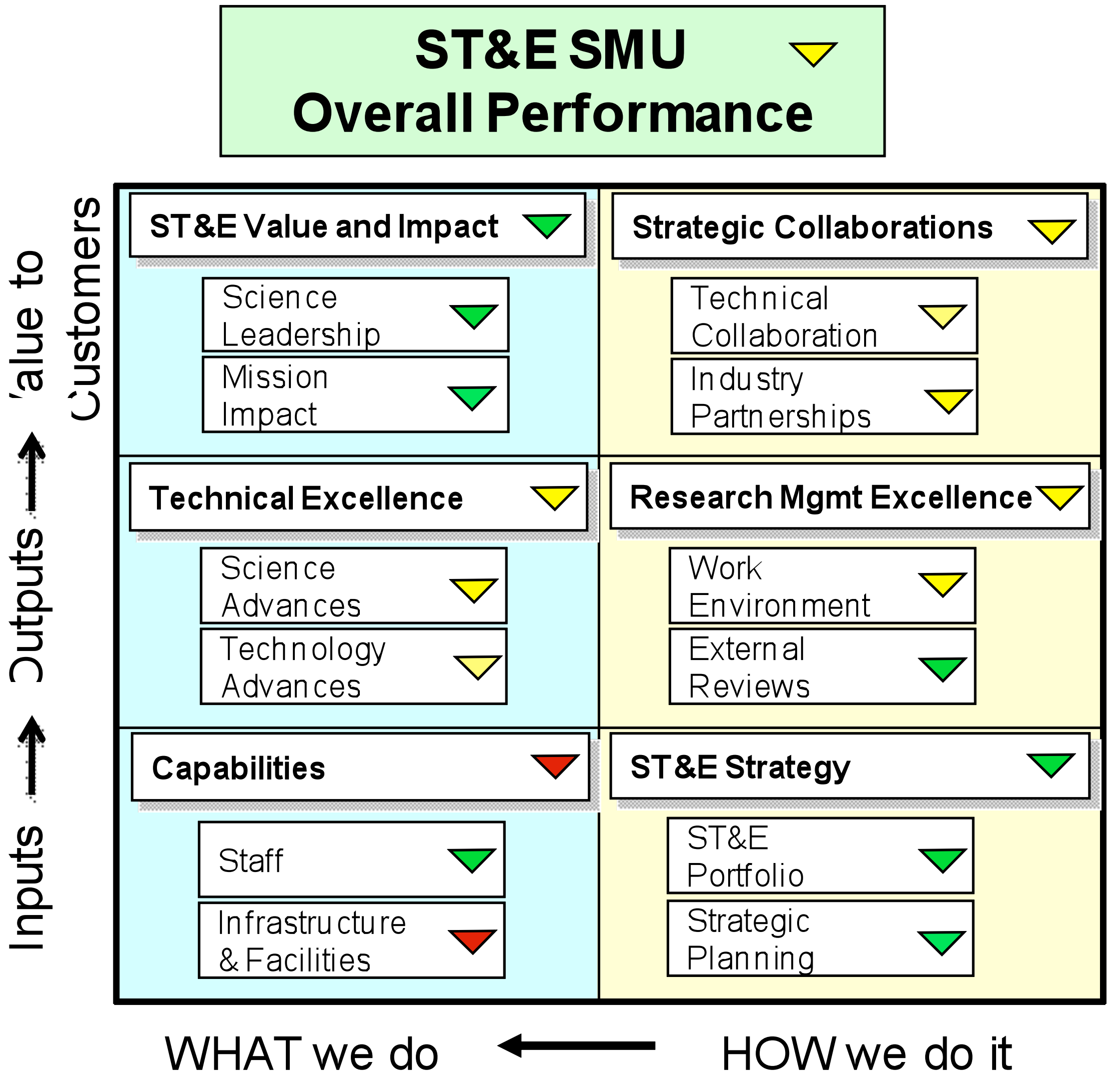


Figure 3-2. The Initial, Integrated, Balanced Set of ST&E Metrics

Trend data is being collected where this is possible. Few benchmarks are available, although there is hope that agreements could be made with sister laboratories to develop these. A system to set targets for the metrics has been defined. Currently, targets are either "maintain steady state" or "increase from previous years." Status categories of "Meets Expectations," "Concern-Watch," "Requires Management Attention," are derived based on the current year as a percentage of previous years, adjusting that percentage by the amount of variation in the historical data. "Quad" charts have been developed for some of the metrics to summarize why management sees the metrics trend to be an issue, and to describe decisions made on what action to take moving forward.

The first Annual Metrics Report was completed in March 2009, in the form of a simple dashboard where it is possible to drill down to the metrics that make up those larger categories, such as "Value and Impact" and "Technical Excellence." The dashboard and metrics are posted on the ST&E SMU web site and updated quarterly with new information. An operating plan has been written to complete the data collection on an annual cycle, standardize data collection and storage, and organize presentation and dissemination of the metrics data. There will be a plan for quality assurance of the metrics data and analysis, and for reviewing the efficacy of the metrics effort in order to continuously improve it.

The metrics have been incorporated into the Laboratories' reporting system, Integrated Laboratory Management System (ILMS), and the yet to be completed PerformanceSoft measurement system. The chosen metrics have been incorporated into the evidence requested in the annual DOE Laboratory Appraisal system, the Performance Evaluation Plan (PEP), and Self Assessment (PEAR).

4. THE CORE METRICS PORTFOLIO AND DASHBOARD

4.1 Defining Critical Success Factors and Metrics

Our general approach was to look at best practices in ST&E metrics and "performance management," which is to choose metrics that matter and relate to the ST&E strategy, that is, the mapping of strategies that underlies a good balanced scorecard. The corporate Sandia metrics effort refers to these as Critical Success Factors (CSFs). Measuring these factors and our success in achieving the ST&E SMU vision will provide the data that management needs to determine if we are "on track," and where improvements are warranted. This small balanced set of metrics is constrained by the intention to implement the process with minimum burden and maximum benefit.

The complexity of the ST&E SMU that we are assessing places a premium on having a rigorous framework with which to define and understand metrics. First we defined Critical Success Factors based upon the ST&E SMU strategic plan. CSFs are necessary elements for organizations to achieve their strategy. Metrics are then logically defined to assess achievement of CSFs.

The Sandia ST&E SMU strategic plan (Sandia 2007)[6] states that the ST&E SMU "will create, integrate and apply capabilities needed to address national security challenges." The two Guiding Principles articulated in the plan can be adopted as CSFs:

- **Nurture the Core**: ensure that the fundamental science and engineering core is vibrant and pushing the forefront of knowledge.
- **Enable the Missions**: apply that science base to enable and deliver effective and innovative solutions to current needs and to anticipate future needs.

Both these factors, which have been combined in the dashboard for simplicity, must be achieved through innovation and within realistic resource constraints of the nation and our ST&E sponsors, partners, and customers. Therefore, we add as a CSF the ST&E SMU corporate Goal 3.0, drive innovation, and connect this to cost:

- **Drive Innovation**. Achieve optimal ST&E value for given "cost": produce breakthrough results through integration, innovation, and working smart, that is, provide optimal value.

For guidance in defining appropriate sub goals for each CSF, we reviewed the ST&E SMU corporate objective**,** "Create breakthrough results through Science and Engineering." The vision of the "Transformed State" after achieving this objective is:

> We are an organization that continuously strives to be at the forefront of science and engineering -- we achieve and practice excellence in science and engineering. We do this through strategic investments in capabilities that yield differentiating strengths in areas such as high performance computing and predictive simulation,

---

6 https://wfsprod01.sandia.gov/groups/srn-uscitizens/documents/document/wfs431138.pdf

> microsystems, and large-scale environmental testing. We fully embed computational simulation in all life cycle engineering activities. We accelerate discovery and innovation through strategic partnerships with industry and universities that integrate world-class science and engineering to create breakthrough results for our mission needs.

Guidance was also derived from the criteria the Department of Energy uses to assess ST&E in the annual appraisal of performance. We wanted to be able to map the critical success factors and metrics to the DOE criteria. These four criteria are Quality of science, technology and engineering; Programmatic performance, management, and planning; Relevance to national needs and agency missions; and Performance in the technical development and operation of major facilities where that applies.

4.2 A Portfolio of Metrics in a Dashboard

As discussed in Section 2, we chose the strategy map/balanced scorecard (BSC) as the means to define and display a portfolio of metrics because it is widely recognized as a good method and Sandia has experience with it. Sandia corporate management and the ST&E SMU both have applied BSCs in the past that were abandoned for different reasons – new external requirements, insufficient use and usefulness, personnel changes. Several organizations in Sandia use BSC now, including Center 1700 and Division 10000.

There are usually four perspectives covered in a balanced scorecard. We initially named these Value to the nation/missions (replacing Financial revenue in the private sector), Value to individual customers, Technical and operational excellence, and Capabilities and learning environment. For each of these perspectives we defined measures initially in three areas: Nurture core ST&E, Enable the Mission, and Drive Innovation (management practices). Later we combined the two Value perspectives, so the ST&E SMU considers customers to be internal Sandia Management Units, external sponsors, partners, the ST&E professional community, and the Nation.

The Nurture the Core and Enable the Missions aspects of performance were combined in order to simplify the dashboard to a two by three matrix. Although we combined core and mission work, we recognize that progress in these types of work occurs on different time frames and is often measured differently. The right hand column of the dashboard contains the management and collaboration aspects of performance. If this dashboard were for a private company, aspects of strategy related to revenue generation would be on the left, and those related to cost and productivity would be on the right.

Figure 3.2 shows the dashboard and displays the categories of metrics. The color of the symbol attached to a metric indicates the status of performance on the metric. Section 6.1 explains how status is determined. The performance status in Figure 3.2 are from an early assessment of trends in the data. The rows are the three BSC perspectives of Learning and Growth (Inputs), Excellence in Technical and Managerial Operations (Outputs), and Value to Customer and National Mission (Outcomes). The columns are the strategic objectives on the left side and the management actions to drive toward those on the right side. The resulting six areas are all Critical Success Factors for ST&E at the Laboratories.

These metrics in the dashboard form an **integrated** portfolio of metrics. Achieving one metric is dependent on achieving several other metrics. If expected funds were not received, outputs will not be produced. Or program characteristics affect level of outputs. For example, outputs such as publications depend on the type of work and customer requirements. Sponsors of basic research value publications. However, classified work has security constraints, and analysis of publications must not penalize classified work for not publishing.

4.3 The Specific Metrics

Table 4.1 shows the specific metrics that are being collected within this dashboard. These metrics were defined to be compatible with the Federal **SMART** metrics criteria (Project SMART, 2008), that is, metrics that are **S**pecific, **M**easurable, **A**ttainable, **R**ealistic, and **T**imely. Our overall deployment of the balanced scorecard is also compatible with the general recommendations of the recently completed Sandia corporate performance measurement study (Frost, 2007).

Table 4-1. ST&E SMU and (selected) LDRD Metrics

Value to Customers ← Outputs ← Inputs

Table 1. STE SMU and (selected) LDRD Metrics	
ST&E Value and Impact	**Strategic Collaborations**
ST&E Leadership and Stewardship - World citation factor by key discipline - Number of professional society positions held - Number of professional society awards - Number of invited talks - Number, characteristics of domestic and foreign visitors to SNL **Mission impact** -ratings and milestones met by STE reported by other SMUs in DOE annual report (PER) -Accomplishments	**Technical** - Number, percent of papers that are co-authored internally, and with external co-authors -Presentations and Conference attendance - Number and dollars of University Contracts **Industry** - Number of active CRADAs - Amount of $ in CRADAs
Technical Excellence	**Research Management Excellence**
Science Advances - Number of Peer-reviewed publications, SNL total; And by RF and by discipline within RF - Number of SAND reports, classified, unclassified - DOE PEP rating for S&T (PO-5) - LDRD Publications **Technical Advances** - Intellectual Property: Number of technical disclosures, patent applications, patents issued - Number of licenses awarded - Number of R&D 100 awards Lab, LDRD	**Work Environment** - Overall rating of the research environment, measured by survey - Benefits of RF External Review, or other self assessment **External Reviews** - Annual number of expert/advisory reviews held - Quality of reviewers - % of expert review panel recommendations acted upon - External customer satisfaction
ST&E Capabilities	**ST&E Strategy**
Staff - Number of new hires by job class, type - Quality: % of Tech Staff with PhDs - Retention: the percentage of staff leaving Sandia **Technology Infrastructure & Facilities** - Capital Equipment Investment, SNL by SMG - Trend in Utilization (Annual dollars spent) by SMG in Computing, MESA, Environmental Test	**Portfolio** - Total SMU costs by RF and other groups - by NW/ITS/LDRD - SMU Annual Costs across B/A/D/O - LDRD projects: % Discover-Create-Prove **Planning** - Number of RFs with current strategic plans - % of Milestones met of ST&E O/G/Ms

WHAT we do ← HOW we do it

5. THE ST&E SMU METRICS PROCESS

5.1 Fit with the Corporate Schemes

Metrics are one part of Sandia ST&E SMU performance assurance, along with other internal assessments and external peer review and advisory boards. Sandia has implemented an Integrated Laboratory Management System (ILMS) that necessarily includes performance assurance as well as strategic planning and other management functions. The elements of performance assurance are shown in Figure 5-1, along with a visualization of where the performance assurance process fits in the larger system.

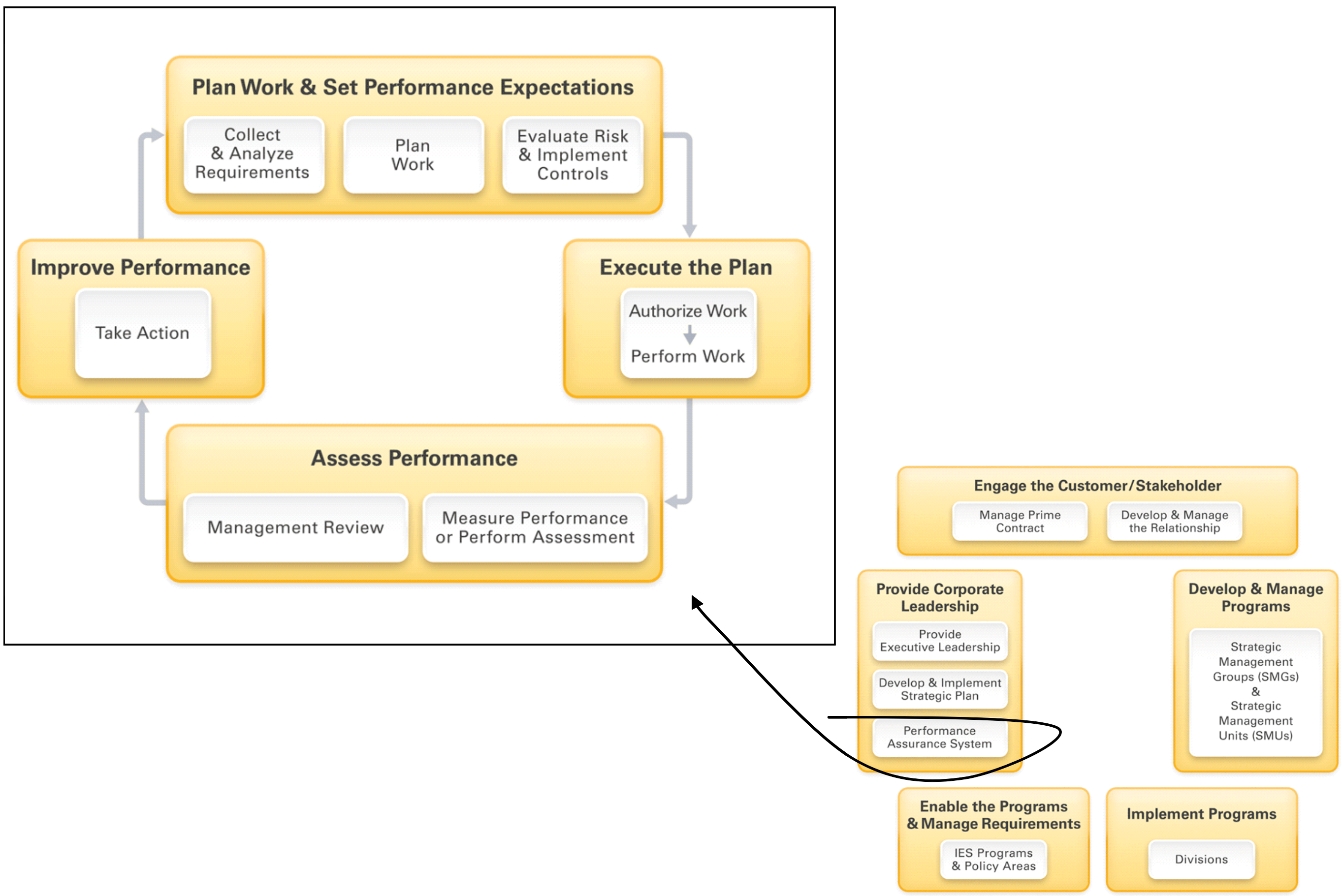


Figure 5-1. Performance Assurance in Sandia's Integrated Laboratory Management System (ILMS)

An essential tool within any management system is the continuous improvement process, by which the management processes and the resulting products are periodically monitored and evaluated for alignment of objectives and results. Sandia uses its Assure, Assess, and Improve Process (AAIP) to evaluate performance against the Strategic Objectives, Goals, and Milestones (OGMs) of the organization. Earlier it was mentioned that the ST&E SMU metrics (Table 4.1) describe performance toward Goal 3.0, Driving Innovation. Thus at the level of the ST&E SMU the corporate process has been replicated. At the corporate level, the AAIP builds on CG100.2, *Develop and Implement the Strategic Plan* and CG100.2.2, *Develop and Maintain Objectives, Goals, and Milestones*, to facilitate prioritizing and executing the essential activities associated with monitoring and improvement of the planning and maintaining of OGMs.

Sandia Laboratories is deploying the PerformanceSoft Views[7] application to address recent performance management assessment findings (Frost Report 2007). The application enables Sandia to capture, summarize and aggregate performance measurement metrics from across the Laboratories. Using PerformanceSoft's data visualization tools (scorecards, "books," drill-down hierarchies), Sandia management should be able to quickly assess the organization's performance.

The plan is for PerformanceSoft Views to become an integral part of ILMS where all can access the information relevant to them. There has been a delay, however, while work is done to improve the selection and visualization of metrics data within ILMS and address concerns expressed in a Value Stream Analysis in March 2009 on Performance Measurement. These concerns included: Complex management and oversight structure drives undesirable complexity in performance management structures. The hierarchy represented by the current, uncoordinated performance management strategy, including structure, alignment, terminology, measures, metrics, OGM, and milestones, isn't clear. Lack of a system approach allows an unhelpful explosion of measures, processes and tools. The current imbalance of qualitative versus quantitative measures and direct/ indirect program emphasis needs correction (Sandia 2009).

The structure for viewing information at the Corporate level currently has ST&E as part of Enable the Mission, alongside Direct the Mission and Corporate Governance," as shown in Figure 5-2.

[7] http://www.actuate.com/products/performancesoft/. Views (PerformanceSoft) is currently in use at Los Alamos National Laboratories.

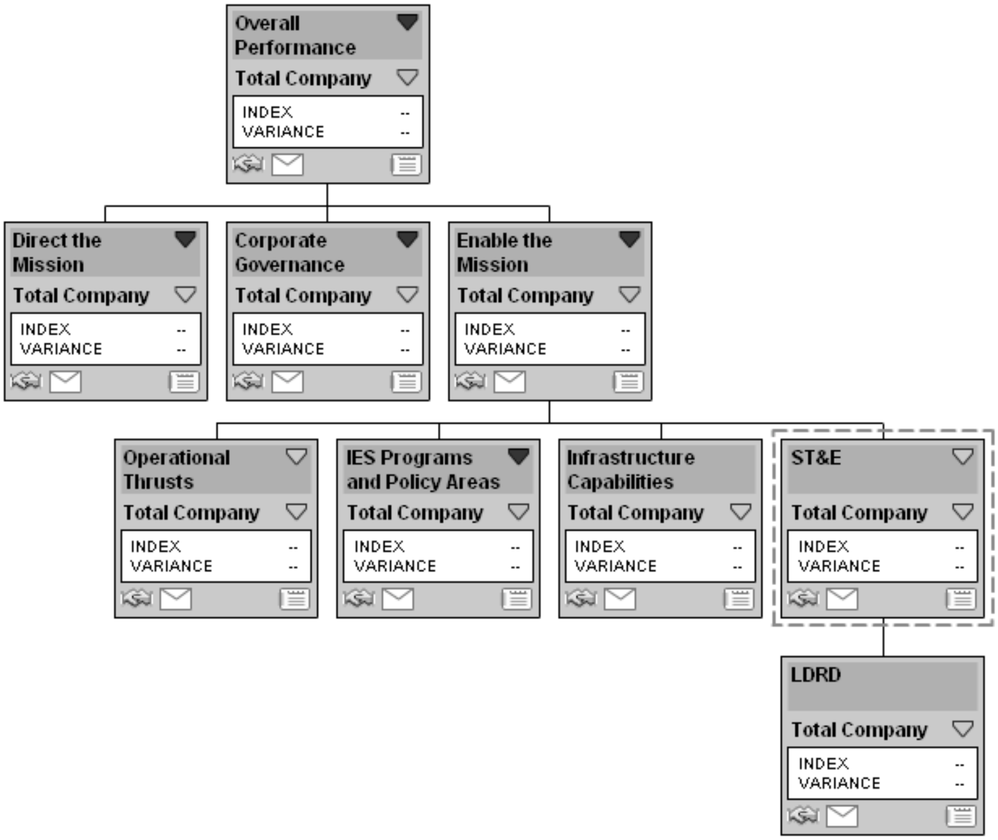


Figure 5-2. ST&E SMU Placement in Sandia PerformanceSoft Scheme

5.2 Fit with Corporate LDRD Metrics

The Sandia LDRD Office has collected quantitative and qualitative metrics data for years. They had developed a scheme for categorizing these metrics just prior to this ST&E SMU metrics effort, and the metrics working group and LDRD Office are still working to reconcile the two sets. The content is very similar, but the frameworks differ. For purposes of PerformanceSoft Views the current scheme includes both ST&E SMU metrics and LDRD metrics.

The LDRD office categorized LDRD program metrics from three sources, (1) Annual Report project summary (submitted by Principal Investigators), (2) Final project review quad chart (submitted by Principal Investigators and Investment Area managers), and (3) completed project metrics (submitted by Principal Investigators). The LDRD program collects metrics on its inputs, processes, project outputs, and program outcomes. Output or outcome metrics were organized into broad outcome categories, and LDRD Investment Areas were provided metrics data for verification and information. The resulting set of metrics is shown in Figure 5-3. Only two of these LDRD metrics are not specifically in the ST&E SMU dashboard shown in Table 4.1. These are “insertion into the current program and "follow-on new projects." These were considered for the ST&E SMU but eliminated to get to a smaller set, with the assumption they could be included in general "accomplishments" metric.

Broad Outcome Category	**Outcome Subcategories**			
S&T Leadership	S&T Advance	External Recognition	Bibliometrics	Professional Awards
Technology Transfer	Patent Application	Patent Issued	Copyright	
Mission Support	S&T Insertion	New Project	New Capability	New Staff
Strategic Partnerships	External Collaboration	CRADA		

Figure 5-3. **Sandia LDRD Impact Metrics Portfolio**

5.3 Overview of the ST&E SMU Metrics Process

The ST&E SMU metric collection, analysis, and communication process supports reporting and management action. The process is a "work in progress" and was designed and will be continuously improved using Guiding Principles that summarize concerns from focus groups (See Appendix B), working group experience, and the literature.

- Operate in an objective manner and always analyze in light of context, so that everyone is treated fairly.
- The process will be transparent to everyone and continuously improved.
- Use multiple measures of value that address both shorter term and longer term factors critical to success.
- Recognize that projects/programs are different and do not each contribute to each metric.
- Maximize the benefits and minimize the costs, both for the overall system and for individual metrics.

Figure 5-4 shows the ST&E SMU process from beginning to end. The intention is to use existing data and data collection processes whenever possible, including data from Human Resources, the Intellectual Property office, and materials prepared for external reviews. ST&E SMU office staff must restructure as well as mine the existing data. For example, reorganizations mean trend data requires mapping people to RFs year by year and full name of authors must be found to display publication data by RF. Doing this provides refined data for the Directors to use for other requirements, such as quarterly reports on high level ST&E risks. When data calls to managers and staff are made, attempts are made to coordinate similar requests and keep the frequency to once per year.

The ST&E SMU Office will develop and maintain a repository for metrics and assessment information thus ensuring some coherence and easy accessibility. The metrics information must be useful and used by Center and RF Directors, the LDRD Office, and ST&E SMU management. Metrics data managed from this perspective provide consistent information for decisions and communication in the Sandia Assurance Process which includes Objectives tracking, DOE Performance Reporting, and ST&E SMU Risk, LDRD, and University assessments. Metrics data can also be used for ST&E SMU Council planning, external advisory and review board discussions, and external communications such as reports to DOE or Congress.

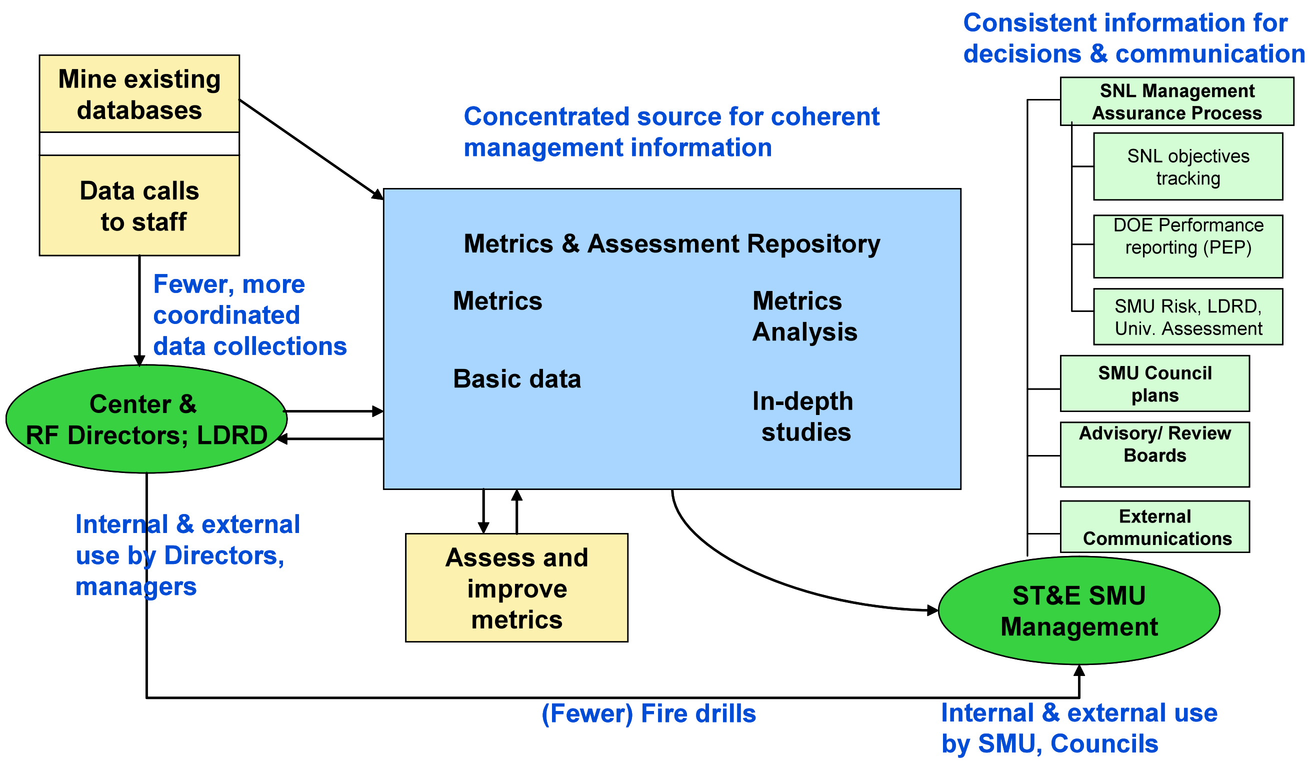


Figure 5-4. ST&E Metric Collection, Analysis, and Communication Process

5.4 The Annual Data Collection Process

We have established an annual process for collecting and analyzing basic metrics data, shown in Figure 5-5, that guides the implementation process described in Figure 5-4. Systematically gathered data on an annual cycle are necessary for sustaining the metrics strategy. The annual cycle reflects the fact that data becomes available at different times of the year, and respects the need for work to be spread across the year to maintain a balanced ST&E SMU staff effort for the metrics process. The data collection drives the analysis, as well as ongoing management requests for information. Basic analysis can and should be performed at the time data is collected and displayed. Periodic reporting will be performed on a regular schedule according to requirements, but at a minimum the metrics dashboard will be updated quarterly. Deeper analysis, and response to varied demands can be done as needed and on a flexible schedule.

It is understood that as concerns arise, data will be collected on additional metrics in these areas. One area already identified is extending basic measurement to the classified mission component of Sandia. Furthermore, as we continue to develop specific definitions for metrics, the intention is to use both simple numbers, such as number of staff with Ph.D.s or number of publications, or more complex metrics such as ratios of publications and number of staff, as well as qualitative data, such as ratings by DOE of ST&E performance in mission areas. The intention is that both quantitative and qualitative data will be as objective, independent, and unbiased as possible. Contrary to what some think, qualitative data, such as might be gathered in a well designed survey process, can be objective and quantitative metrics can be subjective, such as numbers based on self reported data by the persons whose performance is being assessed.

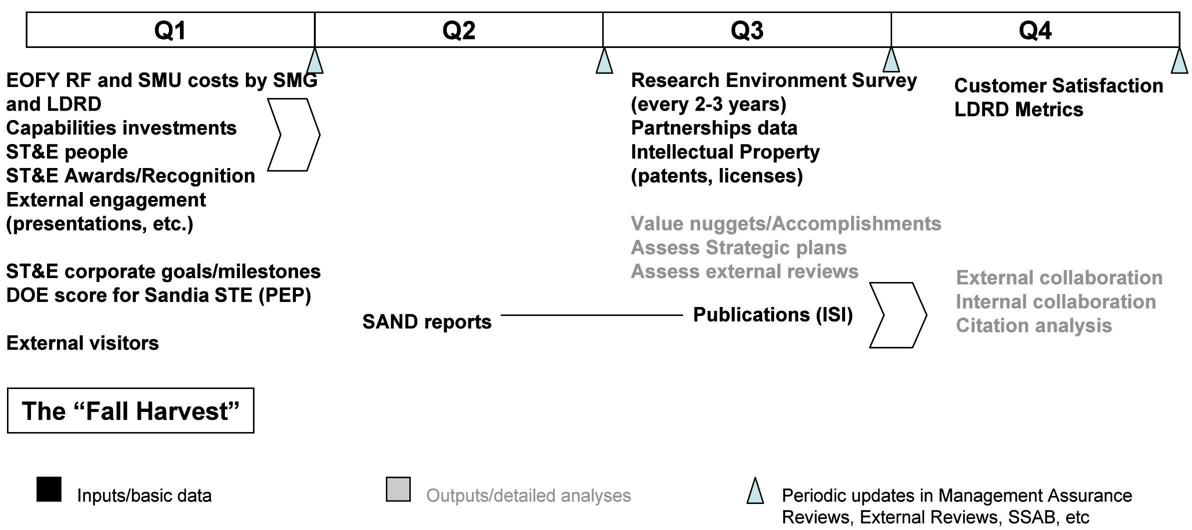


Figure 5-5. Timeline of Annual Data Collection, Analysis, and Reporting

5.5 Data Analysis

The metrics data must always be accompanied by some descriptive analysis of what the data show. The people viewing the metrics data often will not have time to do this analysis. The analysis also ensures that observers all see the same things.

Since we are giving people snapshots, and limiting the amount of data we show, analysis will often add to the metrics information shown. For example, our metric on publications shows data for multiple years by RF. Since we cannot show total publications on the same scale, total number of Sandia publications and how that has changed is noted in the explanation next to the chart. Explanations are also included in the spreadsheet that holds the graphic, and on notes pages of slides.

In some cases there are “nested” sets of data on a single metric where additional disaggregated data is available in slides that “drill down” from SMU to RF, for example, or from total publications to publications by discipline and sub discipline. Or it could be that the sets of data are different aspects of the same data at the same level of aggregation. For example, the metrics data on the ST&E SMU portfolio provide basic data on the total dollars spent that year by RF. The percent of work by categories of Basic, Applied, and Development research is another view of the same dollars.

In other cases analysis may take the form of root cause analysis. Here “digging down” is developing deeper understanding of “Why” the current status of a metric is what it is. The question might be “Are there the right ST&E inputs in place to generate the desired outputs?”

Some basic accepted rules underlying and constraining ST&E SMU metrics data analysis are:

- analysis should be accurate and unbiased, discussing only what can be legitimately concluded from the data;
- analysis must account for normal variance in the data;
- analysis should attempt to separate the signals from the noise, that is, the important from the less important; and
- generally speaking, it takes at least 7 data points to declare that there is a trend.

There is also analysis and interpretation of the metrics data as it is presented to ST&E SMU management for discussion, particularly by those familiar with the data and its context. For example, RF Directors can provide interpretations of any emerging trends in historical data on costs or publications by RFs.

Typical questions to be aware of and discussed when metrics data are reported include:

- Do the data seem accurate and valid to those who know the program well and would use the data?
- What is normal variation in this metric? Is there a trend and if so, what is the trend?
- How does measured performance compare to what we want or expect?
- What other metrics and data shed light here?
- WHY are these data the way they are, that is, what are root causes?

ST&E SMU metrics data analysis has important challenges that influence the type and strength of conclusions that can be drawn. These include:

- Measuring at the appropriate level of aggregation (often the breakdown of data is more useful (and accurate) than gross numbers);

- Prioritizing the information we are working to understand more deeply (e.g., root causes), as well as the questions we are trying to answer, given time and other resource constraints; and
- Understanding sophisticated and coupled information.

5.6 Data Format and Display

Recognizing that these metrics will be shared and potentially viewed by a diverse audience, it is imperative to maintain a consistent presentation format. In this final format, as shown in Figure 5-6, every produced metric should include:

- A graph or graphic depicting the result. Graph types will vary with data types and characteristics of metrics. Graphs should maintain a common look and feel across all metrics (colors, fonts, etc.), and graph types should change across reporting periods only when absolutely necessary.
- The data source, e.g., ThomsonReuters ISI. Include contact information when appropriate.
- The data table from which the graphic was produced, when applicable, e.g., in Microsoft Excel.
- A description of the metric.
- A description of how this metric fits into the PerformanceSoft framework, if applicable.
- The analysis and interpretation of the result.
- A narrative that includes any data cautions or relevant details critical to understanding the metric.

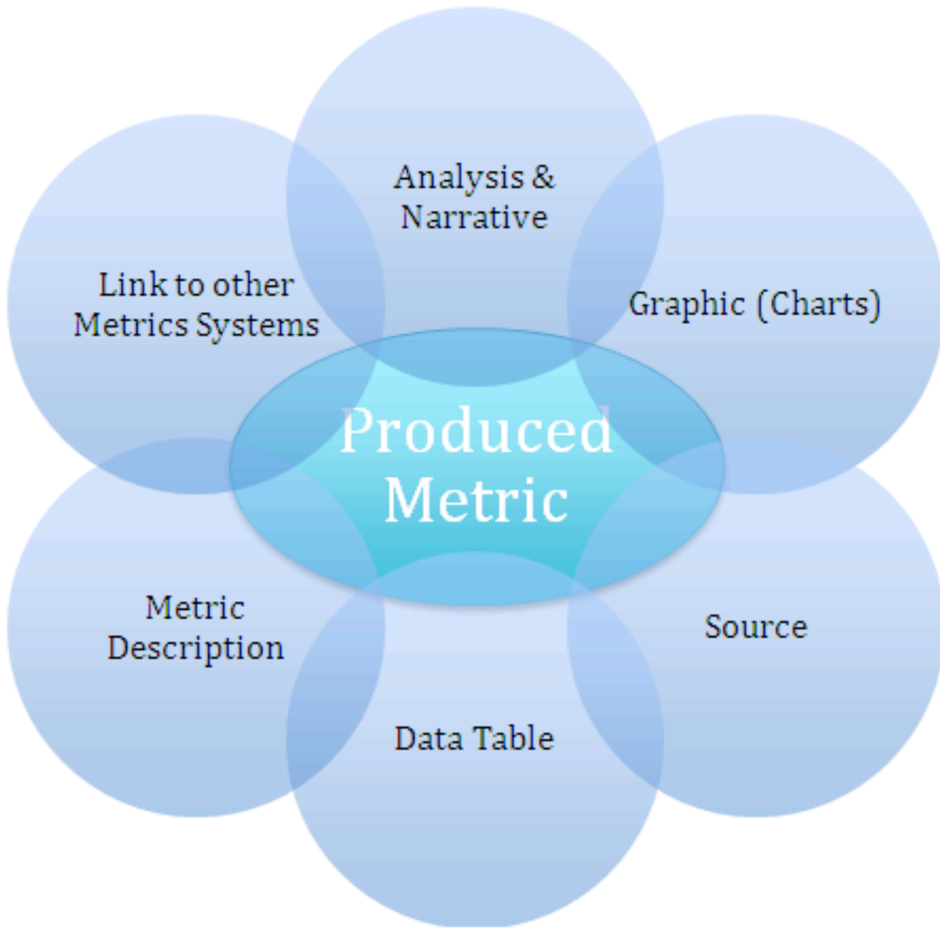


Figure 5-6. Elements of Each Metric When in Final Form

5.7 The Process for Developing Individual Metrics

The first time a metric is collected, analyzed and displayed is a learning process for all of the stakeholders. We have a plan for how this process will be done to ensure that we have quality information in a consistent format so that people accessing the information can interpret and use the information appropriately, as well as easily drop it into a presentation or report. Figure 5-7 depicts processes for both new and mature metrics.

Infrequently, one will find that other groups at the laboratory have applied significant effort examining a set of data and thus the amount of ST&E SMU staff effort required reduces simply to locating these people and their work. More commonly, an analyst will locate raw data and then will be required to transform the data into meaningful information. For instance, ThomsonReuters (formerly ISI) annually provides the Laboratories with an updated database of peer-reviewed publications. However, because of limitations in the data format, publications must first be matched from the database to on-roll staff members at the laboratory. In this case, the metrics team developed custom software tools to merge the annual publication data with human resource data to create a meaningful list of authors and their organization number within an acceptable rate of error. While this is perhaps an extreme example, it illustrates that new (or emerging) metrics sometimes require considerable time and energy during the collection and development stages.
Through experimentation and team evaluation, the development stage of a new metric yields a repeatable process that streamlines work during subsequent reporting periods. A metric is considered mature when the process clearly defines the data source, final presentation format, and outlines the work required for production and analysis of that metric.

The intention is that definitions and the data collection process are clear and standardized. While most metrics will be quantitative, few are simple numbers with simple targets applying to an obvious set of people or work. For example, number of patents granted is influenced by the type of work and corporate patenting strategy. Thus cautions about the data and its use need to be attached to the data. Further, qualitative information such as that from expert judgment, surveys, and interviews, will require accessible links to full reports, rather than just single adjectival ratings or sparse summaries.

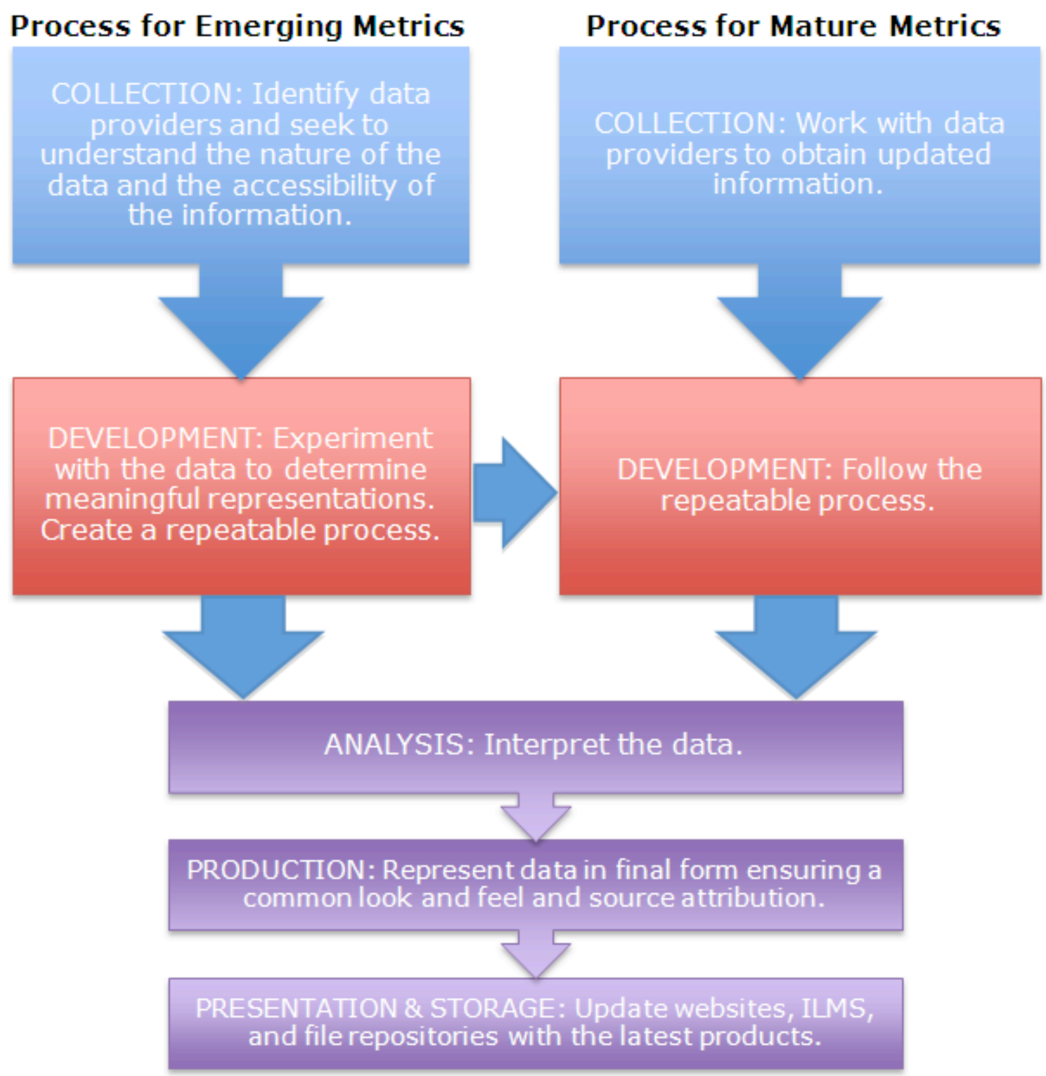


Figure 5-7. Process for Developing New vs. Mature Metrics

5.8 *Continuous Improvement*

Resources are required for the complicated process of collecting, analyzing and acting on a metrics portfolio that is assessing a complex organization. The success of this metrics effort depends on judicious use of resources and balancing the usefulness of data with the burden of collecting it, as well as use of the metrics by all levels of management. Management and metrics staff must resist the temptation to measure everything. Further, expensive in depth analysis must be prioritized. For example, for now, costly retrospective studies of ST&E impact (outcomes that can be attributed to Sandia's work) will not be done.

As the metrics project matures, we expect the costs of ST&E SMU metrics staff time to decrease. Additionally, algorithms are being written that will reduce the person-hours it is taking initially to clean data and organize data, such as binning Center/ department data into Research Foundations. The ST&E SMU repository built for all performance data, including these metrics, and its accessibility to the SMU and SMU Directors, will reduce the fire drill characteristic of data responses to various queries and reports.

Continuous improvement depends upon active participation of stakeholders, both the managers who utilize the data and the people whose performance is being assessed. Sustained commitment of senior managers to developing, using, and improving metrics is essential for success. Input from other stakeholders on how the metrics process is or is not providing benefit to them will help formulate plans for improvements.

There is a vision of the journey to an optimized and sustainable metrics system for the ST&E SMU. Maturation of performance metrics has already proceeded from ad hoc collection for external requests to planned annual data collection and reporting. Optimally this will move to the point where performance metrics are systematic and applied for improvement. This progression is shown in the Metrics Maturity Model in Figure 5-8, developed by Bob Frost of Measurement International.

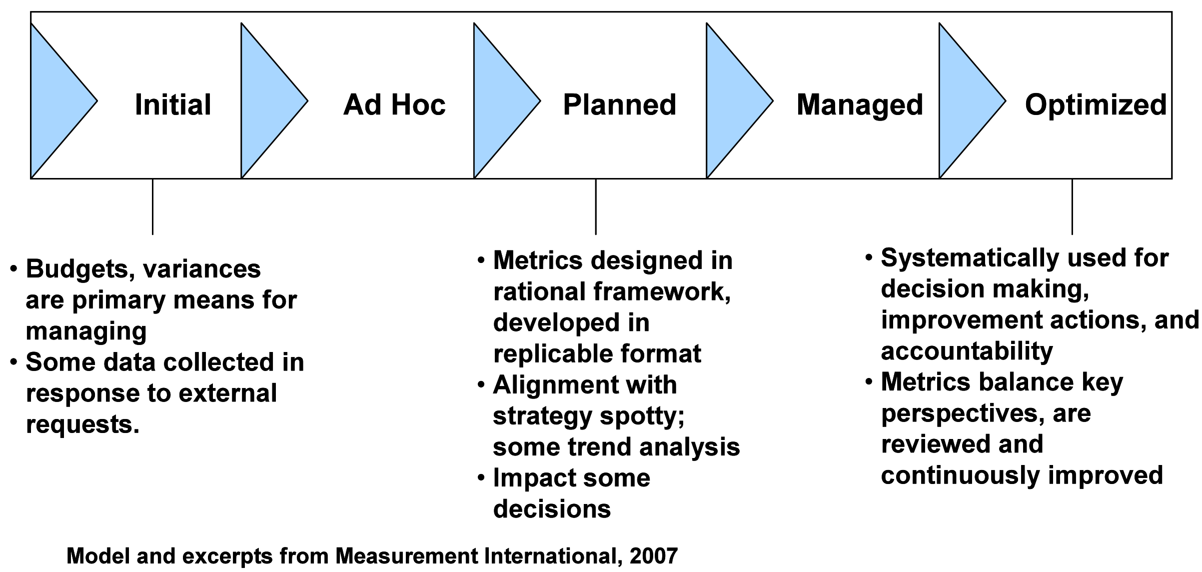


Figure 5-8. Increasing Metrics Maturity

Continuous improvement is one of the key factors in sustainability of a metrics process. To achieve this, we must implement a ST&E SMU metric process that also evaluates the metrics portfolio and implementation process. Factors that contribute to quality improvement of the ST&E SMU metrics process include:

- Plan in partnership with stakeholders;

- Implementation starts small and grows carefully;
- Subject the metrics portfolio and implementation plan to expert review, and revise as needed;
- Devise a lessons-learned component and document the findings; and
- Assess the metrics process annually with participation of various stakeholders and develop an action plan for improvements in next cycle.

Metrics data must be collected and analyzed carefully and we must be mindful that analysis not only responds to questions but will surely generate further questions. Our process must remain flexible and open to excursions in data collection and analysis to meet evolving needs of the ST&E management team.

Current issues we have identified are discussed in the next section. These issues are:

- Utilization of ST&E metrics (setting benchmarks and targets, management use and role in driving behaviors);
- Assuring data quality;
- Adding additional metrics; and
- Improving linkages among metrics.

This page intentionally left blank

6. LOOKING AHEAD

This report documents and communicates the results to date of the metrics project. Our future progress requires that the metrics working group assist with, and track, the application of metrics data and analysis. The group must also complete, monitor, and improve measurement of the initial set of metrics, both individual metrics and integrated analysis, and build a coherent and accessible repository for the metrics. This section looks ahead at our plans for accomplishing these requirements.

6.1 Setting Targets and Using Metrics to Drive Behavior

By definition a metric is a unit of measurement that has a base level defining where you are, a target defining where you want to be, and some notion of the time frame for achieving the target. Setting targets can be better understood with measurement underway and information systematically communicated.

Targets, also called benchmarks, reflect the desired future state for a metric. Thus they are the means of determining whether the measured state of a metric shows progress toward the targeted state and the extent of the remaining gap.

Targets can be determined in several ways, but the ultimate responsibility for setting targets rests with Sandia management. But management needs information on what is feasible and desirable in order to drive desired behaviors. In some cases, there is empirical evidence that more or higher is better, so a general directional trend can be used. Of course, we want to know we aren't near some "maximum or minimum." For example, there is evidence that a higher percent of industrial cost share in a federal technology development project means increased likelihood that ST&E will later be further developed by industry and thus have an impact on products and the national economy. In other cases, an organization may be widely seen as successful, say in building revenue, in part because of its exceptional delivery system, and we could use characteristics of that delivery system as targets for our delivery system if that were important to our success. In yet another case, targets can be determined by looking at what others similar to us in mission space and in this specific metrics area have been able to achieve. For example we could compare our physics publications to those of Los Alamos National Laboratory, although it would be best to normalize the data first for number of Laboratory staff who publish in physics.

Setting targets relates to resource allocation. Success in achieving targets, and goals generally, requires that the system have appropriate resources and incentives in place, of which ST&E SMU metrics are a part.

Until we undertake a more careful search for benchmarks, perhaps by engaging our strategic partners or other laboratories, we are relying on comparisons across time, and across organizational units within the ST&E SMU, and in a few cases with similar external organizations. Our system of scoring performance is based only on comparison across time, with most metrics data defined such that the same or more satisfies expectations and that decline is cause for concern. Scoring is a percentage of the prior year measurement, or percent deviation compared to the average of the past five years. The range of the percentage of prior year measurement is larger if variation is larger. For example, if there is large variation, “Meets expectations” may be “greater than 60 percent of the prior year measurement” rather than the typical 80 percent of prior year. The performance score designations and colors are shown in Figure 6-1. Overall performance is a weighted average of the six measurement areas under it.

- Meets expectations
- Concern-Watch
- Requires management attention
- Insufficient data or targets not established

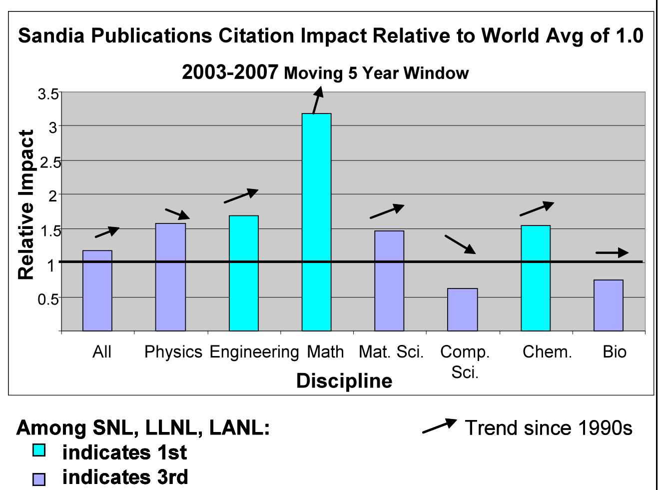
Figure 6-1. Scheme for Scoring Performance

We are proposing that a “Quad chart” be developed after management discussion to summarize the status of a metric, the issues and a path forward, particularly if expectations are not being completely met. An example from actual data is shown in Figure 6-2.

ST&E Value and Impact Citations

A quad chart

Sandia is above the world average in number of citations per paper with two exceptions. Trend appears positive.



Issues:
In key disciplines we don’t compare well to our sister Labs. Looking at Sandia and 6 other labs (ANL, BNL, LANL, LBNL, LLNL, ORNL, SNL) Sandia ranks 1st in Math, 3rd in Chemistry, and 5th in Engineering. In this Lab group Sandia is 7th overall and in the other disciplines.

Source: Thompson Scientific/ISI, 2008 data will be purchased in April/May

Path Forward:
Driving more innovation and more external engagement (current initiatives), as well as encouraging more high quality publications in some areas should improve our ranking over time.

Figure 6-2. Example of a Quad Chart Reporting Metric Status, Path Forward

At the beginning of this effort, the ST&E SMU metrics working group proposed uses of ST&E metrics based on best practices in performance based management. While there has been considerable progress in metrics being used for external review panels and advisory boards at DOE, SMU and RF levels, there needs to be more management discussion on how to integrate these metrics into the strategic management processes and the existing laboratory incentives structures.

The original proposal for management utilization is as follows:

- Selected metrics, where they apply, will be part of the reward system (e.g., PMFs), from VP to staff member;
- Annual consideration of the dashboard by the ST&E Council to be input to Strategic and annual performance plans
    - After consideration by individual research foundations and affected Centers so this can be input for Council
    - Consider along with other data such as Risk Assessment, Customer survey data, and Peer Review reports;
- An annual corporate response and action plan that encompasses all performance information and reporting;
- Annual report to LLT and Sandia Science Advisory Board on accomplishments, value to customers and mission, and capabilities now and for the future; and
- Dashboard available on ST&E home page, with current status of metrics and ability to drill down for more explanation.

6.2 Assuring Data Quality

While we have been conscientious about data quality in the metrics effort to date, the intention is to define and implement a more formal data quality assurance plan during the next year. Poor quality data leads to user/customer dissatisfaction, increased operational cost and less effective decision-making. Since much of the metrics data come from corporate sources, or in the case of publications are purchased from a private firm, this quality assurance plan must extend to the data of others as well as those the ST&E SMU requests or generates. Data cautions will be provided for each metric, such as a very brief summary of expert opinion on the problems with any citation impact data.

We will define data quality broadly. "Data are defined to be of the required quality if they satisfy the requirements stated in a particular specification and the specification correctly reflects the implied needs of the user."[8]

As Marlman points out, Wang, an expert in this field, details a hierarchical approach to data quality based on data quality attributes, dimensions and categories.[9] Data attributes

[8] Marlman, Karen, Sandia, 1999

[9] Richard Y. Wang, Co-Director Total Data Quality Management (TDQM) at MIT, http://web.mit.edu/tdqm/www/contact.shtml

are defined as a set of characteristics required for the objective and measurable assessment of conformance and utility, and hence for the relative assessment of data quality.

Commonly used characteristics include accuracy, completeness, consistency, reliability, timeliness, uniqueness, validity. Data quality dimensions are a set of data quality attributes, and categories are inherent groupings of dimensions. Figure 6-3 shows this hierarchy of Wang's. The four categories of data quality are:

- Intrinsic,
- Contextual,
- Representational, and
- Accessible.

Each of these is defined further by the dimensions below it.

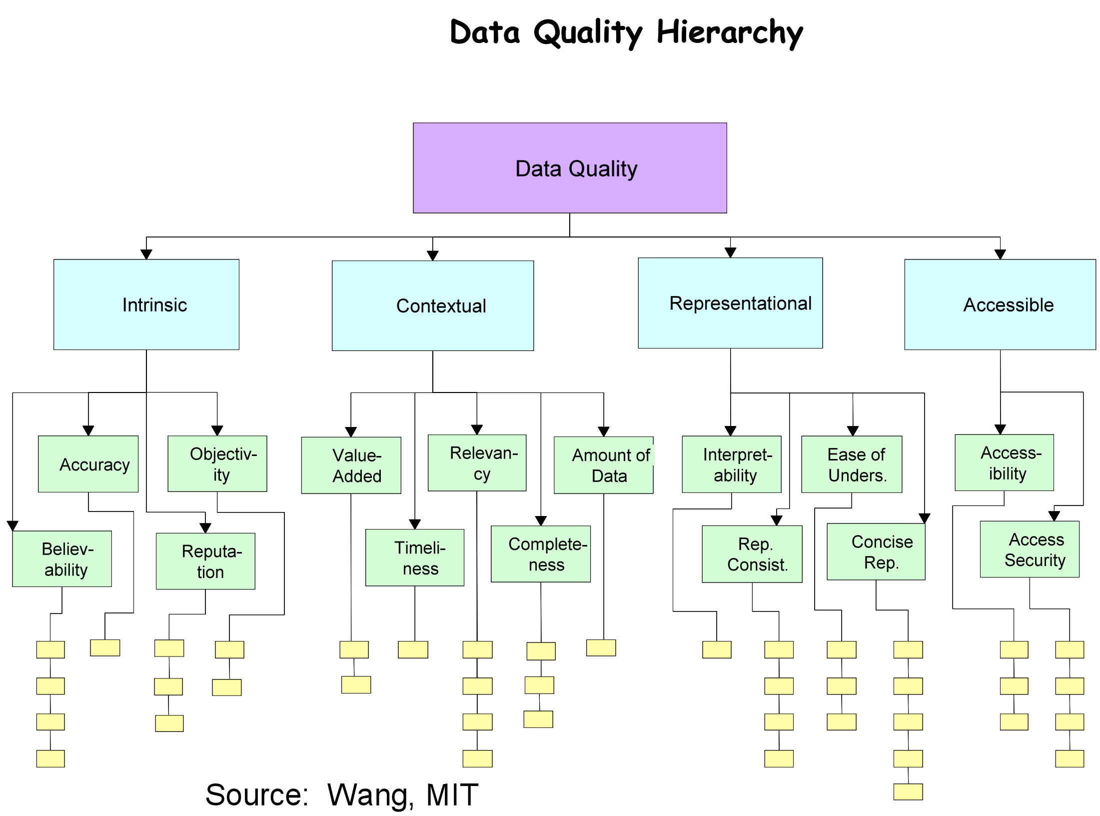


Figure 6-3. A Data Quality Hierarchy

To develop the plan for data quality assurance, we intend to use Wang's hierarchy to (1) identify relevant data quality attributes to the ST&E SMU metrics process, (2) determine our data quality priorities, and (3) assess current data quality. The "fitness for use" contextual dimension will be as important as the intrinsic dimension because we know at the outset that high standards for absolute accuracy are not feasible in the case of ST&E performance metrics.

If assessment shows that data values do not correspond to actual values or are not pertinent to the task of the data user, we will look at data specifications, acquisition of data, and fundamental data processes. If assessment shows that data are not clear or not obtainable by users, we will look at where data does not conform to specifications, the manipulation of data, and technological processes.

6.3 Additional Metrics

6.3.1 Technology Readiness Levels for Technology Programs

An obvious gap in our current set of metrics is a metric that shows progress of the technology development work within the ST&E SMU. Current metrics for progress or outputs (publications) cover the science side of ST&E but not the Technology and Engineering side. One common way of communicating the current status, and thus over time progress, of technology development or engineering is Technology Readiness Levels (TRLs). Readiness levels move through prototypes of increasing scale, where each prototype has measures of technology performance/ functionality and cost in a specific use and intended operating environment. Rather than try to follow specifics such as number of prototypes developed and tested successfully and changes in performance and cost, the intention is that Technical Advances in "Enabling the Mission" will be measured in part by discussing progress through the Technology Readiness Levels shown in Figure 6-3. How many technologies, and/or what percent of technologies, have moved from one stage to another, by technical area? TRLs are used by Department of Defense, NASA, and other organizations. However, not enough Sandia organizations use TRLs for us to have this as a metric at this time.

TRL1 Basic Research
"Experimental data revealing useful information about the basic principles observed"
TRL2 Applied Research
"Model that explains the underlying science and how it could be applied to solve a particular application's problem"
TRL3 Research Result
"Experimental or analytical demonstration that shows that the key elements of an approach are likely to be feasible"
TRL4 Research Demonstration (Lab demo)
"Experiment in a Laboratory"
TRL5 Research Prototype (Demo Unit)
"Looks like a Product, Hand-built by PhDs, Breaks a Lot"
TRL6 Engineering Prototype (Alpha Unit)
"Research Prototype that is Rugged and Repeatable"
TRL7 Flight / Field Prototype (Beta Unit)
"Engineering Prototype that is Reliable and Manufacturable"
TRL8 WR / Hi-Rel (Production Unit)
"Field Prototype that has cost "wrung out" (if applicable), and has completed qualification"
TRL9 Proven Product
"Product that has been used successfully in a system before and is being adapted for use in a similar application.

Source: http://trl.sandia.gov/test_forms_app.swf

Figure 6-4. Definitions of Technology Readiness Levels (TRLs)

6.3.2 Exploring Use of Data Mining to Display Relevance

We are exploring the use of LDRDView, a text-mining tool that is an application of work in Informatics in Center 1400, to describe overlap of Sandia University contracts and Programs to show relevance of the University research to the Sandia mission. SSAB requested that we look at the relevance of university contracts and publications, in addition to counting these. Relevance in this case is the overlap of the subjects being researched in current Sandia contracts with Universities and descriptions of Sandia program missions. Potentially this data mining tool could help us look at relevance of publications and even statements of accomplishments.

ST&E SMU metrics staff entered the statements of work of all active University contracts and narratives describing programs received from Mission program managers into the LDRDView tool. The LDRDView Tool has a text analysis engine that identifies "concepts" (using syntax rules to form chunks of text that reflect content) contained in documents and determines how similar they are to one another. It displays nodes with high similarity values close to each other. The tool allows you to zoom in and out to see different levels of detail, all the way to a specific contract's title, statement of work, and originating Center.

Figure 6-5 is an example of the output of this tool. Programs are yellow squares and contracts are pink squares. It shows in this case that there is little overlap with NW programs, considerable overlap with Homeland Security & Defense (HSD) and Defense Systems and Analysis (DSA, upper right), and some overlap with the more science-oriented areas. A higher level of detail shows that much of the funded University work related to HSD and DSA is in the area of computer and information systems. Work in the area of energy security is also near to the HSD and DSA program areas. In even finer detail, lines show the relationships between contracts and programs, including contract-to-contract and program-to-program overlaps. Each contract has five nearest neighbors and lines drawn between these. The strength of these relationships differs and the tool allows you to examine these strengths and to set the default level of strength to appear on the screen.

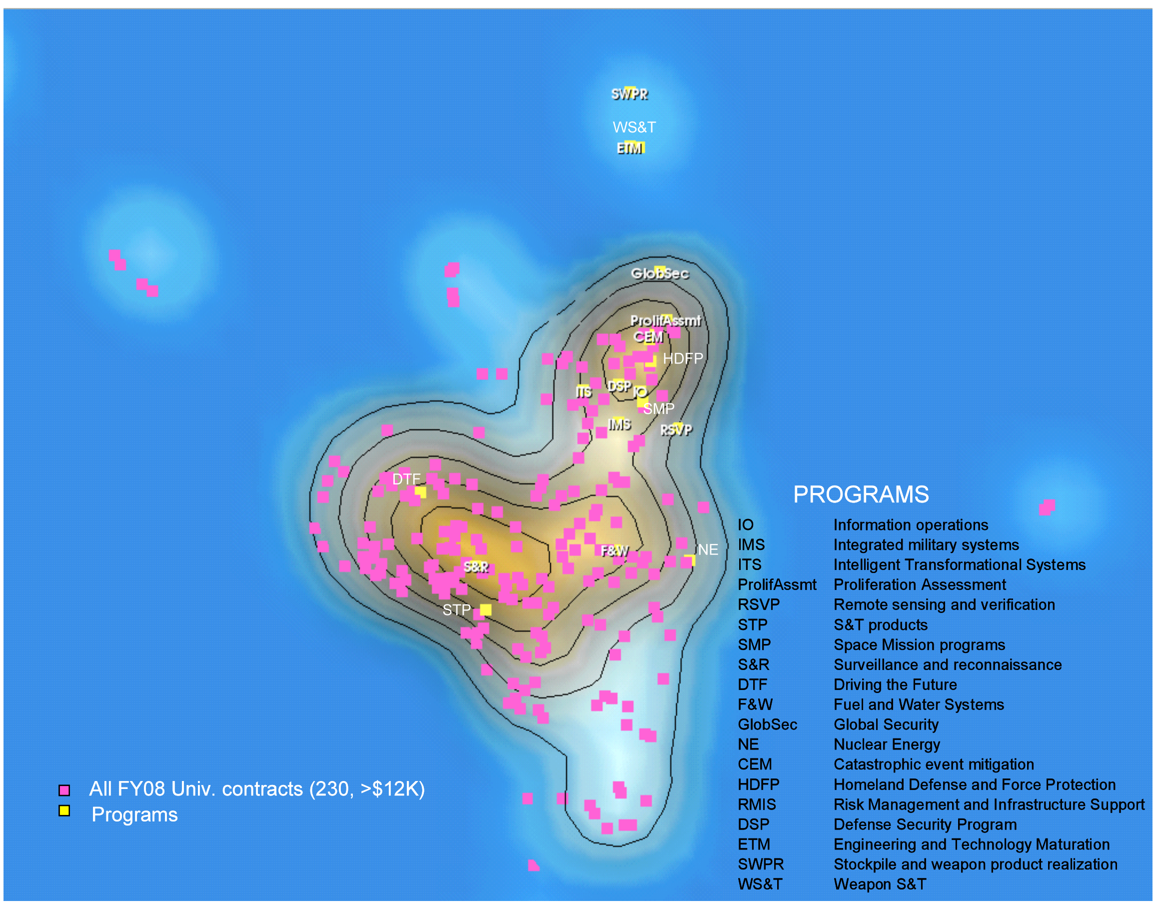


Figure 6-5. Mapping Similarities of Content in Sandia Programs and University Contracts

6.3.3 Improving Ways to Collect information on Value and Impact

Measurements of value and impact are difficult for ST&E programs. We would like to improve on the current situation of using anecdotes and an occasional expensive and limited in-depth study of the impact of a body of research. There are two possible approaches. One is to develop quantitative indicators of customer impact. The other is to improve on qualitative assessments.

A simple leading quantitative metric related to value to customers is assessing whether SMUs, such as NW, or corporate thrust areas have generated and communicated a time-phased roadmap for their technology needs, spanning both near term and very long term, so that it bounds the time horizon for all things "technology". These could be product technologies, assessment technologies, or process technologies. Such roadmaps, if well-communicated, guide technology developments that are linked to longer range corporate obligations. Having metrics around effective communication of technology needs could be an indicator of success in delivering technologies in the future that allow those customers to succeed. More structure could be added to this by identifying TRL levels to be achieved by a certain date, and in some quantitative way, illustrating progress toward those levels. This is complicated by the fact that TRL levels are not determined independent from application, so they may not be adequate as isolated technology metrics.

A possible lagging indicator could be the number of "breakthrough" system / component concepts that have occurred solely through the innovative thinking of the Sandia S&TE community. These aren't necessarily things that the customer or user knew to ask for. They might not have even fit their paradigm of a system or component solution, but when the "breakthrough" occurred, they enabled folks to engineer different solutions, ones that offered advantages that had not been anticipated. A related metric used by industry, though not necessarily requiring a revolutionary breakthrough, is "percent of R&D/ST&E embedded in a final product."

As for more qualitative metrics, rather than in-depth studies such as those explained in the DOE Overview of R&D Evaluation Methods (2007), the plan is to improve on anecdotes. Anecdotes can be collected from lists of Awards, *Annual Lab Accomplishments*, *Sandia Annual Report, Sandia Technology* (a quarterly collection of Sandia Lab News Stories), the LDRD Brochure (from LDRD Day), *Science Matters,* briefing materials from President and Vice President All Hands Sessions, and materials prepared for Expert Review, performance review and Center websites.

We have proposed supplementing these sources for anecdotes, and completed one pilot, of "value nuggets," adapted from a successful multi-year effort of the DOE Office of Basic Energy Sciences. These could possibly be collected at the Department Manager level annually. Staff could use material submitted for many other purposes, and the LDRDView tool could be used for summary display of the information.

A value nugget is a short statement (150-300 words) written for the informed lay person that:

- Summarizes what has been accomplished;
- states the significance of this accomplishment in terms of “change in the state of the art and/ or centrality to field or problem solution;
- says how this has been gainfully used and by whom (in particular for NW, NSTS and external customers); and
- how this can be used in the future and by whom.

There are both science nuggets and technology/engineering nuggets. Value of science advances (nurture the core) includes value of knowledge or research tool or technique to the ST&E community and/or value to business units. Value of technology and engineering advances (enable the mission) includes value of a product (widget or component for widget, algorithm, software, engineering approach, etc.) to the business unit or external customer.

Here is an example of a Value Nugget, excerpted from a two-page *Science Matters* piece and edited by a knowledgeable staff member.

> **Improving Resource Allocation on Supercomputers**
> In collaboration with researchers from the State University of New York-Stony Brook and the University of Illinois-Urbana, Sandia has developed an innovative solution to resource allocation for parallel processing on supercomputers, the Compute Process Allocator (CPA). In experiments, the optimized node allocation strategy employed by CPA increased throughput by 23 percent, in effect, processing five jobs in the time it normally takes to process four. For its superior strategy and scalability over other allocators, the CPA won a prestigious 2006 R&D 100 Award. The CPA's innovative solution was carried to the commercial sector in 2005 when CPA was licensed to Cray Inc. The breadth of impact has been extended through software licensing to numerous laboratory and research centers that bought XT3 systems from Cray.

6.4 Linking Inputs, Outputs, and Value to Customers

If metrics data are truly to influence decisions and improve performance, the feedback loop that connects inputs (staff, technical infrastructure) to outputs (knowledge and products) to outcomes (value to customers and the Nation) must be present and utilized. Current metrics for impact are lacking, and the view at a corporate level connects inputs to outputs to value to customers and the Nation at only a very gross level. This link could be done more easily for a thematic research area, such as Microsystems and Engineering Sciences Applications (MESA), and smaller organizational units, such as RFs or Centers.

For example, for MESA Figure 6-6 shows a measurement strategy that makes the linkages between expenditures, categories of capabilities, products and utilization, and impact. This is drawn from MESA documents, but is a schematic developed by ST&E SMU metrics staff. Figure 6-7 is this same scheme showing more detail.

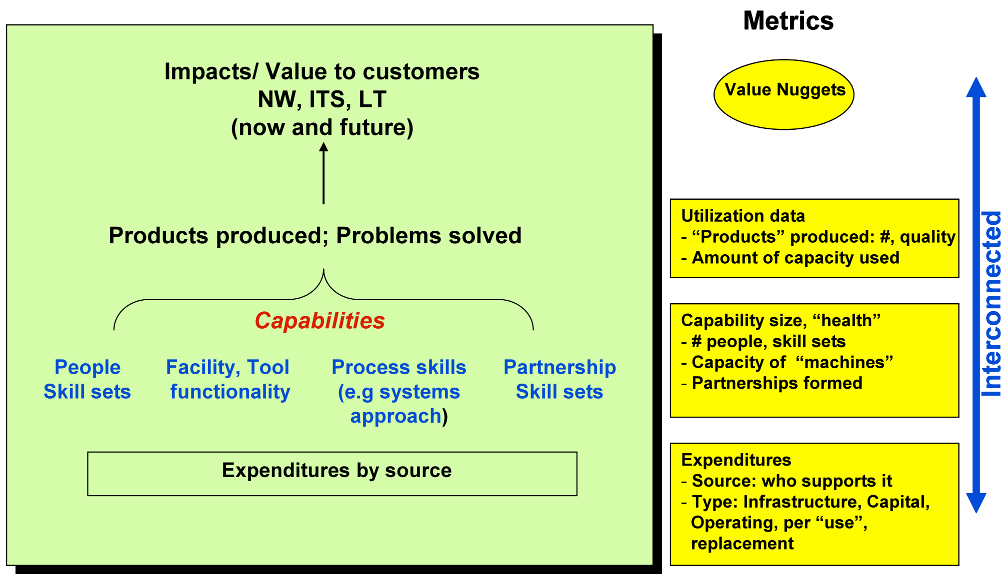


Figure 6-6. Scheme for Interconnected Metrics

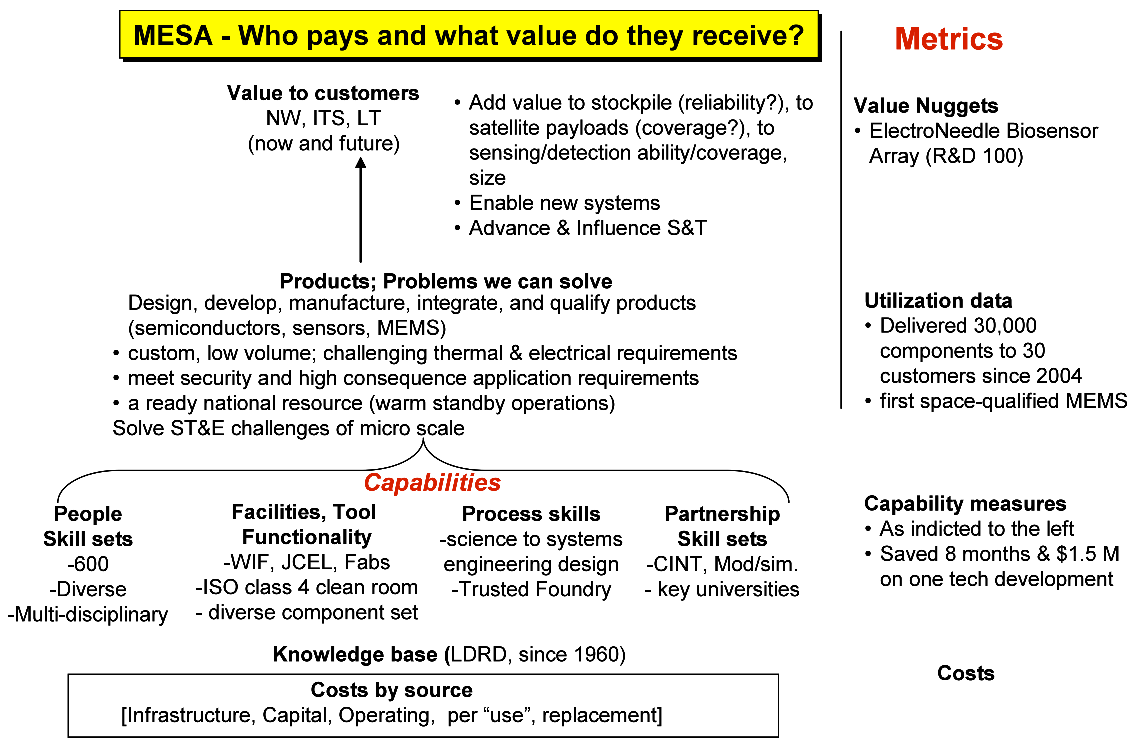


Figure 6-7. Example of Interconnected Metrics

7. CONCLUDING REMARKS

This report provided an overview of why the ST&E SMU established this strategic measurement dashboard and metrics process. Section 2 provided some background on metrics for ST&E programs from existing literature and past Sandia efforts related to metrics, looking for best practices such as balanced scorecards. Sections 3, 4 and 5 provided a summary of work completed to date, specifics on the portfolio of metrics that have been chosen and the implementation process that has been followed. Section 6 looked forward to plans for the coming year to improve the ST&E SMU metrics process.

This is a work in progress. The ST&E SMU metrics team remains committed to helping establish a metrics process that is useful and used for management decisions and demonstrating the value of ST&E here at Sandia. Comments and suggestions for improvements are welcomed.

This page intentionally left blank

8. REFERENCES

Bremser, W.G. and Barsky, N.P. 2004. Utilizing the balanced scorecard for R&D performance measurement. R&D Management. 34 (3): 229-238

Brown, E. A. 1997. "Measuring Performance at the Army Research Laboratory: The Performance Evaluation Construct," The Journal of Technology Transfer, vol. 22, pp. 21-26.

Brown, M. G. and R. A. Svenson. 1998, "Measuring R&D Productivity," Research Technology Management, Volume 41, Number 6, 30-35.

Committee on Science, Engineering, and Public Policy (COSEPUP) (1999), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, Washington, DC: National Academy Press.

Cooper, Robert G. (2008), "Perspective: The Stage-Gate Idea-to-Launch Process-Update, What's New and NexGen Systems," J. Product Innovation Management, Volume 25,No 3, May 2008, pp 213–232

Cozzens, Susan E. (1999). Are New Accountability Rules Bad for Science? *Issues in Science and Technology* Summer:59-66.

Cummings, J., R. Frost, G. Jordan, R. Williams (2001), "Sandia 2001 S&T Benchmarking Study," unpublished internal Sandia National Laboratories document.

Eliat, H., Golany, B., and Shtub, A. 2005. Constructing and evaluating balanced portfolios of R&D projects with interactions: a DEA based methodology. European Journal of Operational Research, In Press.

Ellis, L. 1997. Evaluation of R&D Processes: Effectiveness Through Measurements. Boston, MA: Artech House.

Feller, I. 2003. The academic policy analyst as reporter: the who, what and how of evaluating science and technology programs, in *Learning From Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Phillip Shapira and Stefan Kuhlmann Editors, Edward Elgar Publishing Limited.

Frost, Robert, 2001, "Organizational Metrics Assessment – Sandia National Laboratories," Measurement International Report.

Geisler, E. 2000, *The Metrics of Science and Technology*, Westport, CT: Quorum Books.

Jordan, G.B., J.Mote and J.T.Hage, 2006. "The Balanced Scorecard for R&D: Incorporating Measures of Technical Progress," presented at the Atlanta Conference on Science and Technology Policy, Atlanta, GA, May 18-19.

Kaplan, R. S. and D. R Norton Kaplan, R. S. and D. P. Norton. 1996. The Balanced Scorecard: Translating Strategy into Action, Boston: Harvard Business School Press.

Kaplan, R. S. and D. R Norton. 2000, "Having Trouble with Your Strategy? Then Map It," Harvard Business Review, September-October, 167-176.

Los Alamos National Laboratory (2006), "Metrics for Science and Engineering at Los Alamos National Laboratory," Draft Discussion Document (Official Use Only).

Marlman, Karen S. 1999, Data Quality Improvement Guidelines, Presented to

Federal Aviation Administration (FAA) Air Transportation Oversight System Data Evaluation Program Managers, Sandia National Laboratories.

Modell, S. 2004. Performance measurement myths in the public sector: A research note. Financial Accountability & Management. 20(1): 39-55

Mote, J., G. Jordan and J. Hage. 2007. Measuring Radical Innovation in Real Time. International Journal of Technology, Policy and Management, 7 (4): 355-377.

National Research Council (2007), *A Strategy for Assessing Science: Behavioral and Social Research on Aging*, National Academy Press, Washington, D.C.

National Science Foundation, Science and Engineering Indicators, National Science Board. http://www.nsf.gov/sbe/srs/seind00/start.htm

National Science and Technology Council, Committee on Fundamental Science. (1996). Assessing Fundamental Science. Washington, DC: National Science Foundation.

OECD (Organization for Economic Co-operation and Development). 1997. *In Search of Results: Performance Management Practices.*

OECD (2003), *Frascati Manual 2002: Proposed Standard Practice for Surveys on Research and Experimental Development*, 6th Edition, OECD Publishing.

OECD (2005), *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, 3rd Edition, OECD Publishing.

PerformanceSoft software, see at http://www.actuate.com/products/performancesoft/

Project Smart accessed 2/2008, SMART Goals at: http://www.projectsmart.co.uk/smart-goals.html

Rip, A. 2003. Societal Challenges to Evaluation, in *Learning From Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Phillip Shapira and Stefan Kuhlmann Editors, Edward Elgar Publishing Limited.

Sandia National Laboratories 2008, "Directed Research in Private Industry: How Corporate R&D Drives Innovation." SAND-2008-3809 (Official Use Only).

Sandia National Laboratories 2007, "FY07 Strategic Plan: Science, Technology and Engineering Senior Management Unit," Sandia National Laboratories Report SAND2007-0536P, Printed February 2007 (Official Use Only).

Sandia National Laboratories, 2009, Performance Measurement & Tracking Value Stream Analysis, VSA Out Brief, March 11.

Sandia 2001 S&T Benchmarking Study, 2002, by the Sandia 2001 S&T Benchmarking Team: John Cummings, Bob Frost, Gretchen Jordan, Bobbie Williams, unpublished.

Schay, Barbara and Jerry Robinson, 2008, Directed Research in Private Industry: How Corporate R&D Drives Innovation, Strategic Perspectives, Inc.; Henry R. Westrich (SAND-2008-3809), Official Use Only, June.

U. S. Department of Energy, 2007. Overview of Evaluation Methods for R&D Programs: A Directory of Evaluation Methods Relevant to Technology Development Programs, March 2007. http://www1.eere.energy.gov/ba/pba/performance_evaluation.html

Tipping, J. W. and E. Zeffren, and A. R. Fusfeld. 1995. "Assessing the Value of Your Technology," Research Technology Management, vol. 38, pp. 22-39. See also http://www.iriinc.org/web/.

Wang, Richard Y., Co-Director Total Data Quality Management (TDQM) at MIT, http://web.mit.edu/tdqm/www/contact.shtml

This page intentionally left blank

APPENDIX A -- DETAIL ON ST&E SMU METRICS

Metric	Definition and Data/Data Series	Period & Trend Representation
ST&E Strategy		
ST&E Portfolio Description (funds expended, characteristics of work) B-A-D-O is Basic, Applied, or Development Research and Other (primarily Production)	**SMU Annual Costs (trend)**: 1. Total SMU By RF and other groups 2. by NW/NSTS/LDRD	Annually, starting FY2005; In FY07$$; As % of total SNL $$ From Reportville
	SMU Annual Costs across B/A/D/O 1. Total dollars in SMU, and % allocation of total 2. By RF **LDRD project characteristics** 4. % Discover-Create-Prove 5. S&T challenges 6. Project size	Trends Annually, starting FY07; In FY07 $$ Small data call Annual LDRD reports
ST&E Planning	**SMU Annual Planning milestones** 1. Number of RFs with current strategic plans 2. % of Milestones met of ST&E O/G/Ms	Annual, could report quarterly SMU office tracking
ST&E Capabilities		
Staff (hiring, retention)	**SMU Trends in Hiring and Retention** 1. Number of new hires by job class, by center, for total SMU 2. Number of new hires by category (permanent, postdocs, foreign-nationals) in FTE, Interim, LTE) 3. Quality: % of Tech Staff with PhDs 4. Retention: the percentage of staff leaving Sandia 5. % leaving the SMU –Center transferred from and to **LDRD Impact on Staff** -Staff conversions -Post docs supported	Annual, starting 2004 Compare to size of total SNL workforce From HR Queries LDRD Office tracking
Technology Infrastructure & Facilities (Trends in	**Capital Equipment Investment**: 1. Total $$ invested by NW and ITS	Annually starting FY03
	2. Trend in Annual dollars utilized/spent by SMG in	Annual starting FY06

Metric	Definition and Data/Data Series	Period & Trend Representation
investment generally and for key capabilities)	Computing	Constant FY07 dollars Capital expenditures not included
	3. Annual dollars spent by SMG in MESA	
	4. Annual dollars spent SMG in Environmental Test	
Research Management Excellence		
Work Environment (Research Environment, etc.)	**Health of the Research Environment,** trends for STE SMU as a whole Overall rating of the research environment, measured by survey **Other self assessments**	Survey every two-three years; Have 2008, 2003, 2001
External Reviews (Seek and act upon advice of external experts)	**External SMU, RF reviews** 1. Annual number of expert/advisory reviews held 2. Quality of reviewers 3. % of expert review panel recommendations acted upon **LDRD reviews** 4. NNSA Program Review and Grand Challenge EABs **External Customer Satisfaction**	Annual, starting with 2007 STE SMU office tracking SNL survey
Strategic Collaborations		
Technical collaborations – internal and external	**Co-authorship for STE SMU, by RF** 1. Number, percent of papers that are co-authored internally 2. Number of papers with external co-authors	Annually, starting with FY 2004 STE SMU Office analysis of pubs data
	Trends in University collaborations (Sandia-wide): 1. Number of University Contracts that apply to ST&E 2. Amount of funds to Universities	Annual starting in 2002 From University Office Annual Report
	Presentations & conference attendance	Expense report data
Industry Partnerships	**Trends in Cooperative Agreements (Sandia-wide)_** 1. Number of active CRADAs 2. Amount of $ in CRADAs	Annually, starting FY 2003
Technical Excellence		

Metric	Definition and Data/Data Series	Period & Trend Representation
Science Advances	**Trends in Publication of Research:** 1. Number of Peer-reviewed publications total ST&E SMU, % of SNL total a) by RF and by discipline within RF 2. Number of SAND reports, total and by SMG, classified vs. unclassified **Summary External Judgment on Quality Relevance & Management** **-** PEP rating for PO-5 **LDRD Publications**	Annually starting in 2004
Technical Advances	**Trends in Intellectual Property:** 1. Number of technical disclosures 2. Number of patent applications 3. Number of patents issued 4. Number of licenses awarded **R&D 100 awards Lab-wide and LDRD**	Annually starting in 2003
ST&E Value and Impact		
ST&E Leadership and Stewardship	**Positions and Awards:** 1. Number of professional society positions held by ST&E staff 2. Number of professional society awards each year to ST&E staff 3. Number of invited talks **Citations**: 1. World citation factor by key discipline (physics, engineering, chemistry, bio,) 2. Citations per paper? **Visitors to the Laboratory (ST&E SMU Organizations:** 1. Number, characteristics of domestic visitors to SNL 2. Number, characteristics of foreign visitors to SNL/	Annually STE SMU data call From Thomson/ISI data From Badge Office
Mission impact	**Impact of STE on mission** PEP milestones met by STE reported by other SMUs. **Accomplishments**	Annual STE SMU tracking

This page intentionally left blank

APPENDIX B -- SUMMARY OF FOCUS GROUP INPUT

Proposed Starter Set of Metrics (currently concentrating on those in bold) 2007

Perspective of measurement	**What we will deliver**		**Enabling Management Tools**
	Nurture the Core	**Enable the Missions**	
Have we contributed to the <u>Missions</u> and Nation?	**Recognized as world class ST&E** -**Recruit/retain trend**	**"go to" lab for ST&E solutions** -Reputation for science-based solutions	**Organizations advocate, invest** -National media mentions -Extent of stakeholder advocacy for ST&E
Have results provided value to <u>customers</u>? (SBUs, DOE, WFO, ST&E Community)	**Uses of advances in knowledge, tools** -**Citations, Awards** -**Value of findings & tools to customers**	**Uses of advances in technology or approach** -External investment (Follow on dollars) -Value of ST&E integrated into SMU product	**ST&E and SMUs/customers are mutually supportive** - #, $ in joint work -**External collaborations, partnerships (Co-authorship)**
What <u>products</u> have we produced? (quantity, quality, cost, timeliness)	**Knowledge discovery, create, prove** -**Publications (#, quality)** -"Breakthroughs"	**Technologies are matured** -Improvements in cost or performance -Value to/Movement through readiness levels	**Portfolio, processes encourage agile investment, discovery, maturation** -**ST&E portfolio characteristics** -Effectiveness of select management initiatives
How capable are we of <u>growth</u>/producing, now and in the future?	**We have the capabilities we need now and are building what we need in the future** -**Current capabilities (staff, facilities)** -Progress in building new ones, filling gaps		**Processes, culture encourage learning, taking risk** -**Innovativeness of research environment (employee attitudes on risk taking, etc.)** -Compliance record, burden

Summary of Focus Group Input

Four focus groups were held. The first was in August of 2006 when then CTO Rick Stulen invited comments from the Senior Scientists and Engineers at Sandia on the topic of ST&E measurement. These comments were informally summarized in a memo that provided input and impetus for a serious effort to collect and use ST&E metrics.

During the initial development there were three additional groups of technical staff and lower level managers who participated in focus groups held at both the Sandia New Mexico and California sites in the Spring and Summer of 2007. While basically informal, meetings were structured by a presentation of the proposed strategy and design of the ST&E metrics process, with requested oral feedback from the participants on particular topics that the metrics working group recorded, and the opportunity for further written feedback. The presentation explained the tasking for the metric project from CTO Stulen, the underlying strategy of the Balanced Scorecard (BSC) coupled with a Strategy Map for the ST&E SMU, the metrics working group's initial proposals for specific measures, and near term implementation plans.

The ST&E SMU metrics working group requested discussion and comments from the participants on the general issues of tasking and strategy for ST&E metrics, as well as feedback on the specific measures we proposed for populating the BSC at that time. We requested feedback on the perceived value and fidelity of our defined approach, critique of our proposed measures, suggestions for alternative measures, and critique of the planned implementation, in particular our targets for metric data collection and the intended uses of this information.

In the following, feedback from the focus group sessions is loosely summarized in three general topical areas: (1) general comments and concerns about ST&E metrics and the proposed process; (2) specific metrics concerns; (3) alternative or modified metrics of interest.

1. General Comments and Concerns
    - Why are we doing the ST&E metrics project?
        - Has DOE asked us to do this?
        - Are we aiming the process and collected data at DOE (NNSA)? At Sandia as a whole? At VP 1000 only?
        - What's wrong with other metrics processes, such as the measurement undertaken by LDRD?
        - Use of other existing metrics processes implies more consistency, more connection to "Standards." Use what other people are using – why re-invent the wheel? Using what other people use makes it easier to use the data for new proposal writing.
    - How are metrics going to be used?

  - More specifically, the use of metrics must be very carefully fine-tuned. There is great potential for metrics data collection and use driving wrong behavior. There is also potential for the collected data to be misunderstood.
  - There is concern about the use of metrics in Merit Review.
  - There are varied stakeholders – how will their views of the same metric disperse or diverge? The stakeholder heterogeneity will show up at the Strategy Map level too.
  - Metrics can drive risk-aversion.
  - The ST&E metrics focus needs to support lab investment decisions, especially to guide disinvestment.
  - Metrics hide what is right.
  - Metrics reveal what is wrong.
  - How do we enforce consistency with external measures like Technology Readiness Levels (TRLs)?
  - How will these metrics help anticipate new “market” needs?
  - How do we clarify which projects are expected to produce refereed publications, which should produce follow-on funds, or something similar?
  - How we should best map metrics to the individual, project, program, and overall lab levels is a critical issue.
- Metrics must have a clearly articulated need/purpose:
  - Implication: really use the information.
  - Implication: unambiguous measurement. More generally, we can’t be vague about any aspects of metrics.
  - Implication: the metric must be meaningful within the SNL culture.
  - What are metrics that can be “standard” across the lab, for example across both sites?
- We need both lagging and leading metrics.
  - Accurate metrics are usually lagging metrics, especially if the value of R&D is viewed as the impact and value of its outputs. Time lag in measurement must be understood.
- Metrics need to distinguish what we control versus what we can’t control. That is, “world class science” is judged by others, not by us.
- ST&E metrics highlight the conflict between science and engineering at Sandia.
  - In particular, how can one set of metrics be uniformly applied across such a heterogeneous environment?

  - The difficulty of measuring the "R[esearch]" side and the "D[evelopment]" side makes it hard to find metrics that do both, or that are relevant for an "R&D" lab rather than just a "R" lab or just a "D" lab.
  - Metrics are more likely to reveal that it is a miracle that we do any good ST&E.
  - At SNL, it is "The Core" VERSUS "The Mission" – by definition. What measurement is appropriate for this conflict?
- Sandia's work is unique – therefore what metrics could ever be sensible?
  - Along these lines, by the way, how do you "uniquely measure people?"
  - If a benchmark doesn't exist, how sensible is the measurement?
  - How do we define relevant metrics for classified work?
  - We have many constraints on how we intersect the external industrial marketplace – how do metrics reflect these constraints?
- What is the appropriate time frame for measurement? Time frames can be too short as well as too long. How do we intend to figure this out?
  - We need to extract trends from metrics, which also implies gathering data over a period of time before conclusions are drawn and so on.
- Just because something can be measured doesn't mean it should be measured.
  - Who measures metrics? Only SNL people? Or "anonymous [scary] external people" [who don't understand us]? Or both?
  - Anonymous, external metrics ensure staff cynicism.
- Are we measuring ST&E or are we measuring solutions? The purpose of ST&E is to provide solutions [especially at a lab like Sandia] – ST&E can't be measured independent of this dimension.
- There is the threat that the metric *process* will overwhelm the *product* of the ST&E Metric effort.
  - There is the danger of the metric becoming the goal, rather than the outcome [of using the metric data].
  - The "story" underlying the ST&E metrics is as important as the numbers.
  - The threat is that we will end up measuring behavior more than outcomes.
  - How do we ensure uniformity of implementation?
- We need to get a lot of feedback on the metric process:
  - People in the Tech Staff need to provide feedback on the ST&E Metrics.
  - People at Center level need to review ST&E metrics
  - We need to involve people who can review the appropriateness of the business purpose underlying the metrics.

- Miscellaneous, ungrouped comments.
    - Metrics need to tell you where you go in the next five years.
    - This is just one more thing that takes away from the “real work.”
    - This will cost a lot; it is an incredible time sink.
    - Has SNL spent time with R&D firms that look like us and tried defining/implementing ST&E metrics?
    - [Sandia] response [to incoming work] is often so fast that there is no production of the artifacts that allow measurement.
    - Metrics will fail to influence management because of the second-class status of science at SNL.
    - Metrics drive [or support] an external focus, not internal.
    - The right metrics are just right.
    - Research is heavily government subsidized, therefore not a free market. [Metrics] will therefore be skewed politically. How do we intend to correct for such a bias?
    - At Sandia, compliance is the only thing that matters anymore – why bother with ST&E metrics?
    - Often, research goes off on a tangent – the original goals are not met but something else develops. Is that failure or success? What’s the metric?
    - People find loopholes in metrics; they learn how to “game” metrics; they learn how to use, abuse, ignore, go around and generally destroy metrics processes. What do we intend to do about that?
    - Metrics suggest a standard we can't hope to achieve.
    - Internally reported metrics are as dangerous as externally reported metrics.
    - Good ideas languish and bad ones flourish. Since this firmly rests upon the “personality” of the management team at all levels and in all locations, explain how metrics help address this problem?

2. Specific Metrics Concerns
    - Concerns about the “set” of metrics
        - The Strategy Map looks somewhat internalized – it raises questions about how it interfaces to external customers.
        - Also, can individual staff members see what they are doing through the strategy map and metrics it induces? Are they capable of production/growth? Do they provide products? What value do they provide to customers? How have they contributed to Missions?
        - Ten metrics for the Portfolio [in the 2007 timeframe] are too many.
        - The strategy map and resulting metrics are “overkill.”

  - Why do we have to reflect whatever complexity is being represented in the strategy map?
  - We can't be good at everything. Pick what we want to be good at and apply metrics there.
  - We only need to be great at a few things, good at some things, and get by on most things. Tell us again what your metrics have to do with this realization?
  - Breadth is more important than depth at SNL. How can you measure this characteristic?
  - Continuous improvement is more important than breakthroughs, particularly if you have tight cycles. Measure that instead.
- Measuring publication is important; measuring citation impact is awful.
  - How do we measure publication(s) and impact on the classified side?
  - It is tough to judge instantaneously the value of a publication at the time you submit it. This takes several years. Whether or not the published work is important is subjective. You have to be a historian to understand the value of publications.
- What does "competitive value of research findings and data ..." mean and how do you possibly measure it at a place like Sandia?
  - The only metrics that matter measure: (1) Which customers? (2) How many? (3) What money are you making?
  - Follow-on dollars is an important metric.
  - A dangerous metric is counting dollars – people go after money rather than quality.
  - "World-class ST&E" is a very subjective concept.
- How are metrics projected down to the project level?
- There is too much emphasis on "production" in your metric portfolio.
  - How do you measure "return on investment?"
  - It doesn't matter what you have at the "How capable are you of producing..." level of the strategy map. It only matters "what have you produced."
- Patents are a bad measure because the political/administrative elements in getting patents outweigh the actual technical content they might measure. Another way of viewing this is that counting patents really only is a tip of the iceberg issue even as far as understanding what mechanisms drive achieving patents.
- We need to measure "access" to the lab.
- What does "percentage of funds" mean?
- Don't roll out ST&E metrics the way TRLs are being rolled out.

- "SCORECARDS AREN'T METRICS."
- Provide a Web site that communicates all of this.

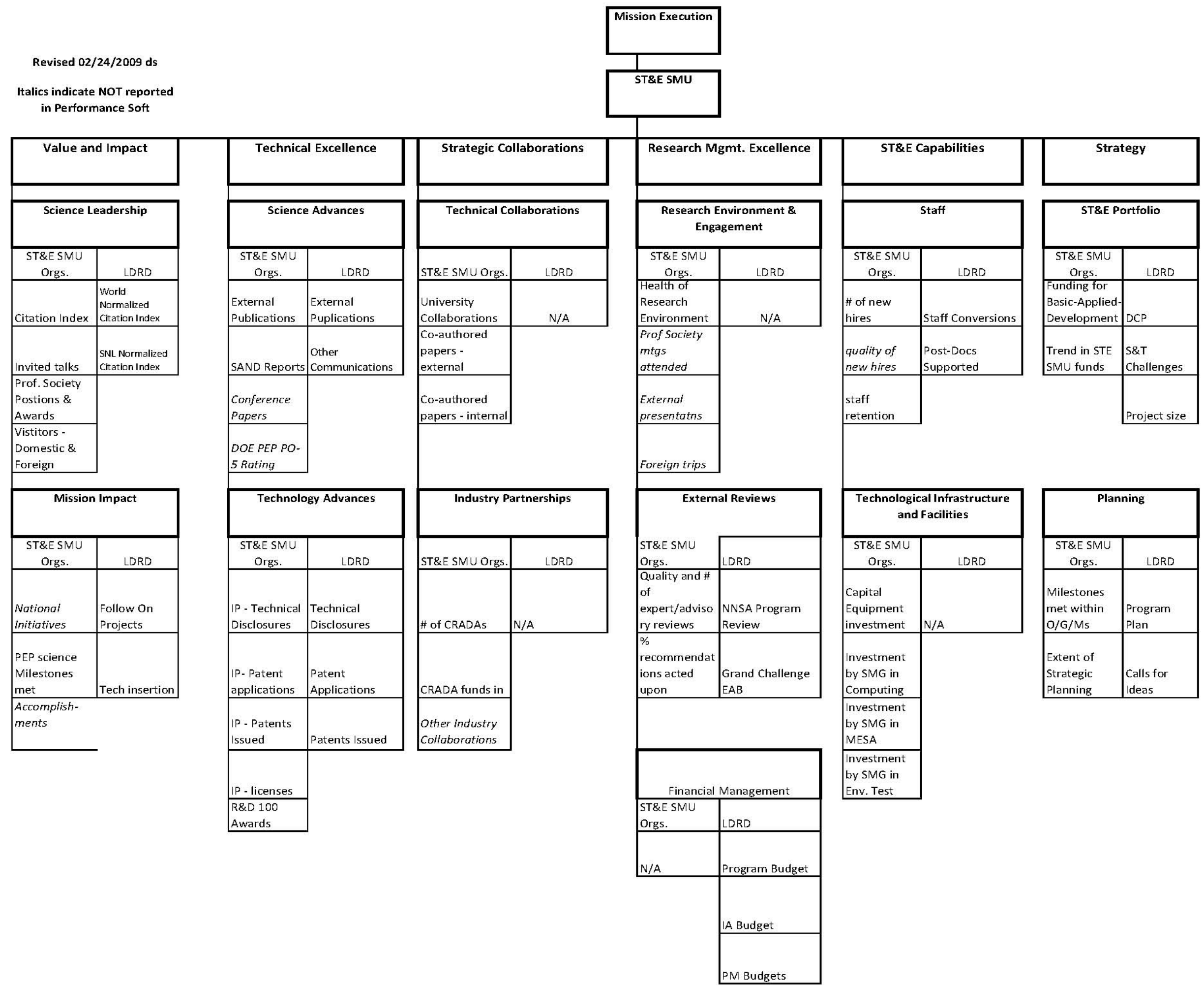
3. Suggested Metrics
   - Dollars the most important measure of "health."
   - Measure the quality, quantity, and frontier nature of capabilities - that is, laboratory capabilities specifically.
     - This gets at measuring capital equipment expenditures, for example.
     - There is belief that our capabilities are good enough for "The [Lab] Mission," but that "The Mission" is a lower bar than "World Class" ST&E. Metrics should clarify what is going on.
     - We should measure the "efficiency" with which we use ST&E dollars.
   - Measure SNL risk-taking.
   - What is the Sandia "Brand" and how can we measure it?
     - Need to measure the connection of ST&E to other business elements within the lab.
     - You should devote energy to measuring all the things done at SNL that are not ST&E.
   - Perhaps some aggregation of the number of papers, citations, and journal impact factors is a metric.
   - A good metric would be one that was centered on how we are replacing "our aging population of researchers" with "talented younger people."
   - Measure both people and program retention.
   - Why are technical researchers still coming to Sandia?
     - You need recruitment metrics.
   - Measure the "stature" of seminar speakers coming to Sandia.
   - You need to measure the Foreign National presence.
     - You need to measure the number of visitors. "If you are great, people come to you."
   - Delivery of a product may be the measure itself - cost and optimal performance may not be relevant.
   - Cycle time in moving through ST&E elements should be measured.
     - Measuring [R&D] execution is critical.
     - Measuring responsiveness is critical.
   - Work on external committees should be a metric.
     - Suggested leading indicator – "focus sessions" at national meetings

- "Technology transfer" and "spin-offs" should be measured.
- Measure why technical staff leaves Sandia.
- Measure short-term changes in Sandia.
- Measure "fragmentation" [that is, project/task fragmentation] in Sandia.
- Measure the difference between "ST&E and paying the bills."
    - How much of a customer's money is spent on writing and going to conferences?
    - One measure would be willingness of [Sandia] SMUs to put money back into ST&E.
- Measure the acquisition of new skills. More generally, measure the intellectual vitality of the lab.
- You need to measure "external visibility."
- You need to measure how technical excellence is recognized within SNL.
- You should measure the amount of funding Sandia is getting in a given research area.
    - You need to distinguish "entitlement funding from DOE" versus funding we've gotten because we are really good.
    - The relationships between funding and people at SNL are crucially important - how do you measure these?
    - This kind of measure (and other measures of external recognition) needs to be sensitive to the size of the external customer base as well as the temporal duration of the funding being measured.
    - You need to measure "prescience" [sic] - how many years back did we do work that led to today's solutions, etc. Are we on the leading edge of waves that are hard to detect?
    - Persistence and history: when did you start investing, how long before first citation?
    - You need to measure program retention.
    - Measure whether customers refer us to other customers.
    - Measure how Sandia is using the ST&E funds. Are we twice as bright as the [a] university? Because we cost twice as much.
- Measure our overhead. It's killing us. Where is the overhead burden coming from? What are people doing with all that overhead money? Etc.
- While you are at it, measure all the factors that lead to "wasted time," not getting work done during the course of the day, week, month, and year.
    - Measure the percent of time a staff member spends in research.

    - How productive is staff when they come to Sandia as compared to before they came to Sandia?
- Measure proposal writing and what comes from it.
- Measure collaboration – is Sandia trying to increase collaboration? Or simply measure it?
    - How do we collaborate? What are the elements of collaboration that can be measured?
    - How about measuring joint proposals?
- We need to do retrospective measurement. Lagging indicators, such as project follow on.
- We need to measure why we fail.

This page intentionally left blank

APPENDIX C -- PERFORMANCESOFT SCHEME (AS OF JUNE 2009)



This page intentionally left blank

APPENDIX D -- DEFINITIONS

Defining the ST&E SMU (Beginning Fiscal Year (BFY) 2010)

In addition to the Research Foundations below, the following organizations are part of the ST&E SMU:
Radiation Sciences (1300)
Manufacturing Process, Science, and Technology (2450)
Energy Resources and Systems Analysis (6300)

Bio Science (BIO):
- 08620 (all departments)
- 08630 (all departments)

Computational and Information Sciences (CIS):
- 01400 (all subgroups)
- 08960 (all departments)

Engineering Sciences (ES):
- 01500 (all subgroups)
- 8249 (multi-Physics Modeling & Sim)
- 8350, 8360 (combustion)

Materials Science and Technology (MST):
- 1110, 1130 (CINT)
- 1800
- 8131 (Rad/Nuc Detection Materials & Analysis)
- 8222 (Hydrogen & Metallurgy Sci), 8223 (Materials Chem.), 8246 (Mechanics of Materials)
- 8650 (Energy Materials Sciences)

Microelectronics and Microsystems (MM):
- 1120
- 1700 (all subgroups)

Pulsed Power (PP):
- 01600 (all subgroups)

Distribution List:

Name, Org.	**MS**	**Name, Org.**	**MS**
Julia Phillips, 01100	1427	Jerry Simmons, 01120	1421
Tom Bickel, 01200	0104	Carol Adkins, 01210	0139
James Lee, 01300	1169	Paul Raglin, 01380	1145
James Peery, 01400	0321	John Mitchiner, 01430	0321
Art Ratzel, 01500	0384	Joel Lash, 01510	0824
Keith Matzen, 01600	1190	John Porter, 01670	1191
Gil Herrera, 01700	1079	Dave Myers, 01770	1071
Duane Dimos, 01800	0887	Justine Johannes, 01810	0885
David Plummer, 02400	0429	Mark Smith, 02450	0961
Terry Michalske, 06300	0721	Pablo Garcia, 06320	1138
Glenn Kubiak, 08600	9405	Grant Heffelfinger, 08630	1413
Bob Carling, 08300	9054	Art Pontau, 08360	9054
Len Napolitano, 08900	9151		
Rob Leland, 09300	0801	John Zepper, 09320	0823
Peter Davies, 12100	0103	Tommy Woodall, 12140	0127
Karl Braithwaite, 12120	0131		
Kent Biringer, 12121	0131		
Hal Morgan, 01030	0351	Mary Monson, 01032	0833
Allen Camp, 00310	0116	Lori Parrott, 12140	0127
Steve Rottler, 01000	0351	Robert Scott, SSO	0184
Charles Barbour, 01010	0351	Dan Sanchez, SSO	0184
Randy Watkins, 01012	0351	Tim Trucano, 01411	0370
Hank Westrich, 01011	0123	Pete Oelschlaeger, 0p1012	0351
Paul Rockett, 01011	9161	Alan Burns, 01012	0351
Marianne Walck, 6700	0701	Gretchen Jordan, 01012	0351
Russ Skocypec, 08004	0484	Dominique Wilson, 01012	0123
Marie L. Garcia, 01012	0123	Jeff Brinker, 01002	1349
Yolanda Moreno, 01012	0123	Carl Petersen, 1515	0825
Nancy Davis, 10610	0351	Laura McNamara, 1433	0370
Joshua Parsons, 10610	0351	Ron Hartwig, 02100	0429
TJ Allard, 09750	1491	Bob McCarty, 12141	0127
Greg Durfee, 09751	1491	Mary Kay Austin, 00215	0136
Deborah Serna, 09751	1491	Mark S. Allen, 01031	0114
Marcey Hoover, 12141	0127	Paul Shoemaker, 00210	0138
Tim Knewitz, 10680	0127	Margaret Lovell, 12200	0185
Jeanne Evans, 00212	0145	Jennifer Nelson, 06060	0724
Paul Yourick, 05210	1231	Chuck Meyers, 09001	0531
Teresa Mills, 106601	0724		
Technical Library, (9536, electronic copy)	0899		